PERFORMANCE-BASED PAYMENTS, PROVIDER MOTIVATION AND QUALITY OF CARE IN AFGHANISTAN

by
Elina M. Dale

Baltimore, Maryland
February 2014

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Abstract

Background: Challenges in attracting new health workers and poor motivation of existing staff have been cited as key reasons for underperformance of health systems. In the summer of 2010, Afghanistan launched a performance-based payments (PBP) program in 11 of its 34 provinces. The goal of this dissertation research is to examine the effect of PBP on health worker motivation and quality of care in Afghanistan.

Methods: In 11 provinces of Afghanistan, 442 facilities, stratified by type of facility and matched on the average number of outpatient visits per month, were randomly assigned to an intervention or a comparison arm. Because of the study design, masking of the participants was not feasible. The main data source for all three papers is a cross-sectional health facility survey conducted 23 months after the start of the program as part of the annual National Health Services Performance Assessment (NHSPA) conducted in all 34 provinces of Afghanistan. In Paper #1, a four-subscale 15-item scale was developed and tested among 2,413 health workers drawn from 775 health facilities across 34 provinces, including Dari and Pashto speakers. Reliability of the scale was assessed using ordinal alpha. Construct validity of the scale was assessed using intention to quit, performance assessment and feedback, staff voice, equity of rewards, and general motivation. The generalizability of the scale was tested on a sample of Pashto speakers. Standard error computations used a sandwich estimator to account for non-independence of observations as a result of cluster sampling. Paper#2 and Paper#3 were limited to the 11 provinces implementing the PBP program. Paper #2 was based on a random sample of 805 health workers from 256 facilities. In Paper #2, the outcome was measured using the four-factor 15 item-scale developed in Paper #1. A latent variable modeling framework

was used to estimate the effect of performance-based payments on four motivation factors. A robust variance estimator was used to adjust for correlation of the data within health facilities. The effect of treatment was estimated using intention-to-treat analysis. Given the high rate of non-compliance, complier-average causal effect approach was also examined. Paper #3 is based on a random sample of 2,180 patients and 255 health workers from 233 facilities. The outcome, quality of care, was measured using a standardized checklist designed to measure clinical quality of care. The main analysis approach was a linear mixed effects model using maximum likelihood estimator.

Results: The work motivation scale developed as part of this thesis demonstrated good psychometric properties. Three out of four subscales had ordinal alpha of 0.70 or higher. The introjection subscale with the lowest internal consistency coefficient among the four subscales (0.64) still demonstrated higher reliability than subscales on comparable measures. The results of the multiple group analysis demonstrating the difference in factor means for amotivation between respondents intending to quit and those intending to stay provided evidence for construct validity of the scale. As hypothesized, staff voice had statistically significant (p-value<0.05) and positive association with identified and introjected regulation factors. Equity of rewards appeared to be a strong predictor of external regulation and introjection (p-value<0.05). Based on the intention-to-treat analysis, no statistically significant changes were observed in health worker motivation in the intervention group as compared to the control group (p-value>0.05). Using a complier-average causal effects approach, it appeared that PBP were negatively associated with motivation factors (p-value<0.05), except amotivation. Comparing patients in the treatment group to patients in the control group the expected change in z-

score for quality of care was 0.23 (p-value<0.05). Other covariates were added to the model. Together, they explained 15% of the total variance in quality of care, while the proportion of level-2 variance explained by the covariates was 26%.

Conclusions: The study produced a valid and reliable measure of worker motivation in Afghanistan. It demonstrated that while PBP program had a significant positive effect on quality of care, there was no effect on motivation. When comparing those health workers who reported having received PBP in the treatment group with similar health workers the control group, the study showed a negative effect of PBP on different types of motivation, including external regulation. At the same time, holding treatment constant, identified regulation was statistically significantly associated with quality of care (p-value<0.05). Thus, it appears that health workers who are more intrinsically motivated do provide higher quality of care, but their motivation was not affected by PBP. Moreover, while treatment, i.e. PBP program participation, appears to be a significant predictor of quality of care, it explains only a small proportion of observed variance in quality of care between as well as within health workers. Factors such as presence of functioning equipment, timely payment of salaries, and type of health worker or level of training seem to be more important predictors of quality of care than PBP program participation, as they explain higher proportion of total variation observed in quality of care. This suggests that PBP should be paired with other complementary quality improvement activities such as training and infrastructure investment.

Thesis Readers:

Professor David H. Peters (Advisor), Department of International Health

Professor Saifuddin Ahmed (Chair), Department of Population, Family and Reproductive Health

Professor Sara Bennett, Department of International Health

Professor Kitty S. Chan, Department of Health Policy and Management

Alternate Thesis Readers:

Professor Qian-Li Xue, Department of Epidemiology

Assistant Scientist Shivam Gupta, Department of International Health

Acknowledgements

This thesis is the result of support and guidance of many individuals and organizations.

First and foremost, I am deeply grateful to Dr. David H. Peters, my advisor, whose knowledge, experience and encouragement made the completion of this thesis possible. I am forever indebted to David for giving me an opportunity to work on the analysis of the Afghan national household survey during my first year at JHSPH – it was an incredible learning experience for me and the beginning of an amazing journey, the result of which is this thesis. It also provided me with financial means without which it would have been difficult for me to continue my studies. I have greatly benefited from David's vast knowledge and practical experience in health systems.

I am also grateful to Dr. Saifuddin Ahmed, the chair of my thesis committee, whose insightful teaching and great patience over the past years were vital for my work in Afghanistan. I knew I could always find help from Dr. Ahmed when I met with a new statistical challenge. I have greatly benefited from rich experience and insights on health policy, human resources and work motivation shared by Dr. Sara Bennett. I am particularly indebted to Dr. Qian-Li Xue whose courses served as the methodological foundation for this thesis. His constant guidance was vital for the completion of this work. I would also like to acknowledge Dr. Kitty Chan for her meticulous, practical and encouraging feedback and comments as I developed the proposal, designed the motivation scale, and wrote the dissertation.

I am grateful to Dr. Joanne Katz, Dr. Marie Diener-West, and Dr. Lawrence Moulton for their insightful teaching and guidance since I started the program. I would also like to acknowledge Dr. Elizabeth Stuart for her helpful comments as I was finalizing my thesis. I am indebted to Dr. Shivam Gupta whose guidance and support were extremely important during my work in Afghanistan. I am thankful to every member of the JHU/IIHMR team. Many thanks go to Dr. Gilbert Burnham.

I would like to thank everyone from the Ministry of Public Health of Afghanistan, particularly the Health Economics and Finance Department. Their hard work and dedication to health reforms in Afghanistan have inspired me; their openness and willingness to share information have been invaluable. My special thanks go to Dr. Najeebullah Hoshang.

I have also learned a lot from other doctoral students. Special thanks go to Aarushi Bhatnagar, Ann Lin, Andrew Mirelman, Kayhan Natiq, Ligia Paina, Pooja Sripad, and Hannah Tappis for their friendship, for sharing their knowledge and experience, and for making the process of writing this dissertation less lonely. I am thankful to Olakunle Alonge for his intellectual support.

I am particularly indebted to Carol Buckley who always had an encouraging word and an open door for me. Carol provided important administrative guidance required for completing my thesis and the degree.

I would also like to acknowledge Dr. Ainura Ibraimova, Dr. Melitta Jakab, Dr. Peyvand Khaleghian and Dr. Christopher Lovelace who first inspired my interest in health systems and encouraged me to pursue a PhD at Hopkins. I am particularly thankful to Dr. Joe Kutzin whose support has helped me to go through some of the most difficult moments of this incredible journey.

I am grateful to my late father, Melis Manzhiev, who taught me to never give up, to work hard, and to pursue my dreams even if no one else believed in them. I hope you can see me now and I make you just a little bit proud. I would also like to thank my sisters Anzhela, Edita, Zita, Rita and Uma whose love, support, and optimism have been invaluable throughout my studies.

Finally, I would like to thank my husband, my friend and my strongest supporter, Ivar Dale. I thank you for sacrificing the first year of our family life together in order to enable me to pursue my dream of getting a PhD at Hopkins. You told me not to give up even when I was ready to do so. You moved half way around the world to be closer to me when I went to Afghanistan. You were always there for me, and we did this together.

Contents

List of figures

List of tables

1 Chapter 1: Introduction and background

1.1 Introduction

After nearly three decades of conflict and the rule by the Taliban regime, Afghanistan faced enormous challenges in education, health, infrastructure and other basic services. In 2001, net school enrollment was estimated at 43% for boys and at 3% for girls [1]. According to the World Bank estimates, there was a desperate shortage of qualified teachers with about 21,000 teachers for a school-age population of more than 5 million (a ratio of 240 students per teacher) [1]. The infant mortality rate (IMR) in 2001 was estimated at 91.6 per 1,000 live births (uncertainty bound 84.9 – 99.3), and the under five mortality rate (U5MR) for the same year was estimated at 131.0 per 1,000 live births (uncertainty bound 120.4 – 143.1)[1] [2]. The country had the highest maternal mortality ratio in the world: In 2000, it had the estimated MMR of 1,957 per 100,000 live births with the uncertainty interval 729–4356 [3]. To address these enormous challenges, beginning in 2002, international donors and non-governmental organizations (NGOs) have provided large amounts of financial and human resources to Afghanistan. To date, the World Bank has committed more than $2.7 billion for development projects in Afghanistan [1].

Since then, the country has made a significant progress in improving access to health and education services leading to improvements in education levels and health outcomes [1, 4-6]. While the exact figures vary by source, it appears that IMR, U5MR and MMR have all declined over the past decade. However, despite this progress according to The State

[1] These estimates are lower than the estimates provided by UNICEF (2012), The State of the World's Children 2012, according to which in 2010 the U5MR was 149 per 1,000 live births.

of the World's Children 2012 report [7], among 193 countries ranked from highest to lowest, Afghanistan is ranked number eleven in under-five mortality rate. It also has the highest MMR of 1,575 maternal deaths per 100 000 live births (uncertainty interval 594–3,396), which is about 394 times higher than Italy, the country with the lowest MMR of 4 (3–5)[2]. Moreover, it is one of the countries with the slowest yearly rate of decline in MMR [3]. Thus, the government and its development partners have been in search of new approaches to improve access to and quality of maternal and child health services.

Results-based financing in health (RBF), specifically performance-based payments (PBP) for health care providers, showed promise as growing evidence from countries such as Rwanda and Haiti seemed to suggest its effectiveness in increasing utilization and quality of health services for women and children [8, 9]. Thus, in the summer of 2010, Afghanistan launched an RBF project in purposefully selected 11 of its 34 provinces. The primary objective of the project was "to increase key maternal and child health outputs, to improve quality of health services and to ensure higher patient satisfaction with health service delivery" [10]. It was hypothesized that improved financial incentives in the form of PBP would improve health worker motivation and satisfaction. This, in turn, would lead to better health system performance, resulting in improved health outcomes. This doctoral dissertation examines the effect of PBP on health worker motivation and quality of care.

[2] The data for MMR are for 2008. These are the latest figures that have been accepted by the government and international stakeholders.

1.2 Literature review

1.2.1 Definitions and measures of motivation

There are a number of different definitions of motivation in general and work motivation specifically, and all of them define it in terms of intent to act or engage in a particular type of behavior, which is separate from the actual performance of the task [11-13]. As Shortell & Kaluzny [11] state, "Motivation should not be confused with performance. People can be highly motivated but still perform poorly." It cannot be assumed that motivated workers are more productive, because performance depends not just on motivation but also on ability and a number of other factors such as the availability of resources needed to perform a job well [14]. Also, while highly interrelated and often used interchangeably, job satisfaction and motivation are distinct constructs [15-17]. Job satisfaction is more affective in nature, and it is expressed in attitudes and emotional state; while motivation is directed by particular goals and is "the reason why workers behave as they do towards achievement of personal and organizational goals"[16].

According to Shortell & Kaluzny [11], and originally proposed by Campbell and Pritchard [13], theories of work motivation can be divided into two basic types of theories: content and process. Content theories focus on different types of needs and desire for satisfaction of these needs as the basis for motivation; whereas process theories attempt to describe how motivation is created and maintained (Table 1.1).

The Cognitive Evaluation Theory (CET) was the first theory to clearly describe motivation as something that can be created by external circumstances but also as

something that arises from within the individual. According to Latham and Pinder [18], work motivation is "a set of energetic forces that originates both within as well as beyond an individual's being, to initiate work-related behavior, and to determine its form, direction, intensity and duration". However, the CET was not well equipped to describe the work motivation, particularly as it relates to financial rewards [19]. It incited a debate on whether and how external rewards in general, but financial rewards in particular, had negative consequences on motivation. According to DeCharms [20], such rewards shifted the locus of causality and people became either controlled by contingencies or de-motivated.

Like CET, the Self-Determination Theory (SDT), described by Deci and Ryan [21], distinguishes between intrinsic motivation (i.e., doing an activity for its own sake) and extrinsic motivation (i.e., doing an activity for an instrumental reason). People do not require external reasons or rewards to do activities that that they find interesting or "aesthetically pleasing", whereas they need extrinsic motivation for doing other activities [21]. However, according to SDT, extrinsic motivation is more complex than this simple definition. There are different types of extrinsic motivation, depending on the extent to which it is controlled by external factors. These types of motivation can be aligned along a continuum (Figure 1.1), "representing the degree to which goals or values have been internalized" [21]. The key distinction between SDT and most other work motivation theories is that it emphasizes the multidimensional aspect of motivation as a construct and thus, focuses on different types of motivation rather than on its total amount [19].

According to Gagne and Deci [19], different types of motivation are associated with different behavioral outcomes.

According to Levesque et al [22], amotivation is the absence of intent to act in a particular way. External regulation describes behavior that is driven only by an expectation of a reward or fear of negative consequences. Introjected regulation is like a bridge between controlled and autonomous motivation. The regulation comes from within but it is controlled by external forces. Examples of this type of regulation include "contingent self-esteem, which pressures people to behave in order to feel worthy, and ego involvement, which pressures people to behave in order to buttress their fragile egos" [19]. Identified regulation describes behavior in which people act because the activity is congruent with their personal goals and beliefs. Gagne and Deci [19] give an example of a nurse who performs uninteresting and not necessarily pleasant tasks such as bathing patients because she values her patients' health and understands the importance of these activities for their well-being. Finally, with integrated regulation, individual intention to act comes from their sense of self, and a given behavior is part of who they feel they are: In the example with the nurse, it means that the nursing profession is central to her identity, and she "would be more likely to act in ways that are consistent with caring for people more generally" [19]. According to Tremblay et al [23], these types of extrinsic motivation are not like developmental stages which everyone has to go through. The type of motivation one has at any given moment depends on particular organizational context and personal experience. As Tremblay et al describe it [23], SDT focuses on the "nature" of motivation, that is, the "why of behavior".

Work motivation is influenced by type of management, system of rewards and other organizational factors, or what is often called organizational structure and culture. Organizational structure is "the formal system of task and authority relationships that control how people coordinate their actions and use resources to achieve organizational goals" [24]. Organizational culture is "the set of shared values and norms that control organizational members' interactions with each other and with suppliers, customers, and other people outside the organization" [24]. Thus, the principal purpose of organizational culture and structure is to shape behavior and to "control the means used to motivate people to achieve [organizational] goals" [24]. Organizations differ in their values and norms as well as formal systems of authority and task relationships, and differences in these can be summarized by differences in organizational design, which is the process through which managers select and use different elements of structure and culture. According to Jones [24], studies have shown that, for example, all else being equal, the more hierarchical the organization is – the less motivated individual workers are, because the relative authority and the area of responsibility decrease.

An important cautionary note is appropriate here regarding the concept of motivation and factors that support its different dimensions. Up and until now the discussion has largely ignored the fact that the motivation theories reviewed here have been developed in cultures that value autonomy and individualism [25]. For example, Jones' [24] assertion that hierarchical organizations are negatively associated with worker motivation may not be applicable to non-Western societies. For example, Schwartz [26] describes settings

where the person is viewed as "embedded in a collectivity" and the feeling of autonomy is not strongly valued. According to Kao and Sek-Hong [25], in the Chinese enterprise, consciousness about moral obligation is central to work motivation and thus, introjected regulation – viewed as a controlled form of motivation far away from intrinsic motivation by Deci and Gagne [19] – may be predominant and considered to be intrinsic in settings where feelings of duty before one's family and community are far more important than the desire for autonomy.

As described above, SDT conceptualizes motivation as a multidimensional construct. Based on this theory, a number of scales have been developed that ask participants why they do particular things. Different types of motivation are reflected in different reasons for behavior, which are organized into different subscales. According to Gagne and Deci [19], based on the research conducted so far in education, sports and health behavior changes, each subscale correlates most positively with the subscales closest to it and negatively with subscales that are conceptually on the other end of the continuum (Figure 1.1), which is consistent with the SDT. Subscales can be used individually to predict outcomes such as performance, or they can be combined algebraically to form a summary measure [27]. The Self-Regulation Questionnaires (SRQ) and Intrinsic Motivation Inventory (IMI) have been developed at the University of Rochester based on these principles [27].

The SRQs are a series of scales originally introduced by Ryan and Connell [28]. The SRQs measure a range of subjects from academic motivation to exercise, but none of

these have been designed or validated in an organizational setting [22, 28]. A Work Extrinsic and Intrinsic Motivation Scale (WEIMS) is a scale based on the SDT but developed specifically to measure work motivation. It has been tested among the Canadian forces [23]. The Motivation at Work Scale is another scale based on SDT to measure work motivation [29]. Unlike WEIMS, it has been tested and validated among a more diverse group that included workers in sectors such as sales and services, technical and manual, managerial and professional, and health and education. It has also been cross-validated in two languages, French and English. However, it is still limited to Canada. Thus, the context for which it was developed and in which it was tested was very different from the focus of this study. As described in the previous section, the cultural context of the theories that serve as a basis for these instruments is very different than the one proposed in this study. Instruments developed and tested in countries such as France, Canada and USA have to be significantly revised to be appropriate for settings such as Afghanistan.

1.2.2 Quality of care theory and measurement

According to the Institute of Medicine [30], quality of care is "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." It has six key characteristics. It is safe, effective, patient-centered, timely, efficient, and equitable [30]. For the assessment of quality of care Donabedian's three-part framework has been widely used [31]. It consists of the following elements [31]:

1. Structure, defined as material and human resources as well as organizational structure of the facility;
2. Process, defined as actions taken by the provider in making a diagnosis or treating the patient; and
3. Outcome, reflecting effects of care on health status, behavior and satisfaction of the individuals and the population.

As Basinga et al [32] noted in their study on PBP in Rwanda, the most significant effects were observed for those services over which providers had the most control. Individual health worker efforts are unlikely to result in substantial changes in structure or outcome measures of quality. Outcome measures, particularly changes in the health status of patients, depend on a complex set of factors, including those outside of health provider's control. The effect of PBP schemes on the structural aspect of quality of care depends on the design of the program. In many PBP schemes, including Rwanda, part of the bonus payments was spent on improving structural aspect of quality of care, such as purchase of new equipment and repairs of the overall facility infrastructure. In Afghanistan, however, while the recipients of PBP were facilities, the entire amount of bonus payments was distributed among individual health workers as salary top-ups. Thus, in Afghanistan no direct investments were made into facility-level improvements from PBP. Also, while a team of motivated workers in an organization could influence structural changes, by for example, introducing innovations to improve drug availability or increase their own qualifications, many aspects of structure typically depend on the managing NGO or regional health administrators. It seemed that if one were to observe any changes as the

result of PBP, they would first of all be in processes of care. Thus, the main focus on this study was on the second element of Donabedian's framework, i.e. performance of the practitioner, specifically technical and interpersonal processes of care [31].

As described in Franco et al [33], there are six main methods for measuring quality of provider performance:

1. Exit interviews with patients or caretakers about provider performance,
2. Standardized simulated patients,
3. Review of medical records,
4. Direct observation of actual care using a structured checklist,
5. Simulation/demonstration, and
6. Clinical vignettes or other types of knowledge assessments.

Review of medical records is probably the least intrusive and expensive method, but it is limited by the quality of records, which maybe particularly weak in developing country settings [33]. Also, many procedures go unrecorded and some measures of quality, such as friendliness of the provider, cannot be recorded at all. Exit interviews measure reliably only certain aspects of process of care, such as interpersonal characteristics, while technical aspects of care might be less well understood, particularly by largely illiterate population [34, 35]. Using standardized simulated patients is resource-intensive and is rarely feasible in large scale evaluations [34]. Also, it is not usable for assessing care for children under-five. Clinical vignettes or other types of knowledge assessments as well as simulations aimed at assessing practical skills are not good measures of performance because studies [36, 37] have shown there is a gap between what health providers can do

and what they actually do. Although providers are likely to modify their behavior under observation [36], direct observation using a checklist is generally considered to be a gold standard for other assessment methods [33], and in the context of developing countries with largely illiterate populations it has demonstrated the highest overall sensitivity and specificity [38]. In Afghanistan, direct observation for measuring quality of care has been used as the gold standard for measuring clinical quality of care, although it is supplemented by exit interviews to provide patient perspective on quality of care [34, 39]. It has been used for several years now as part of the National Health Services Performance Assessment (NHSPA) [40].

1.2.3 Theory and evidence on the impact of performance-based financing on work motivation, utilization and quality of health care services

As Eldridge and Palmer [41] noted in their review there was some inconsistency in the use of terms such as performance-based payments, or pay-for performance. Performance-based payments (PBP), also referred to as performance-based financing (PBF) can be defined as a mechanism "by which health providers are, at least partially, funded on the basis of their performance" [8]. Musgrove provides a more narrow definition and puts PBF in a larger context of results-based financing (RBF). According to the definition provided in Musgrove's glossary of terms [42], RBF refers to any program that rewards the delivery of one or more outputs or outcomes by one or more incentives, financial or otherwise, upon verification that the agreed-upon result has actually been delivered. Incentives may be directed at service providers (supply-side), program beneficiaries (demand-side) or both. PBF is a form of RBF distinguished by three conditions:

- Incentives are directed only at providers, not beneficiaries;
- Awards are purely financial – payment is by fee-for-service for specified services; and
- Payment is made only if those services are of approved quality, as specified by protocols for processes or outcomes.

PBP is often framed in terms of principal-agent theory [43]. A principal-agent relationship arises whenever one actor (the principal), which can be a person or organization, contracts, either explicitly or implicitly, another actor (the agent) to perform services or supply goods. However, because the agent has a utility function that differs from the utility function of the principal, the agent does not always act in the best interest of the principal. Moreover, since the principal cannot always observe the agent, resulting in information asymmetry between the two parties, the issue of moral hazard arises. Thus, the central problem for the principal is how to ensure that the agent acts in the way that is desired by the principal [44]. PBP is seen as one of the approaches to resolving the principal–agent problem by designing such contracts where the agent will be motivated to act in the best interest of the principal because their objectives are fully aligned [41, 43]. This approach is based on the assumption that financial gains are chief motivational forces for the agents [41].

The assumption of overarching importance of financial incentives has not been thoroughly investigated yet in cultures where it is a matter of pride to be a health worker and while it may not be the most financially rewarding profession, it is one of the most

respected ones [45-47]. Social status was a recurring theme in interviews with clinical officers in Tanzania for example, although in this case it seemed that social status was closely connected to salaries because low salaries undermined the status of these health workers in their communities [48]. As Franco et al [33] point out in their study of motivation among health workers in Georgia and Jordan, providing adequate salary is an important factor but there are a number of potential non-financial mechanisms for improving work motivation.

The literature coming from the field of organizational psychology is not as uniform in its opinion of financial incentives as a means of motivating employees and thus, improving their performance. According to the theory of reinforcement or operant conditioning, employees should be rewarded for behaviors that are desirable by their management (reinforcers) and punished, or at least not rewarded, for undesirable behaviors (punishments) [49]. The two main types of reinforcers used to elicit the desirable behavior are financial or verbal. However, in opposition to reinforcement theory, DeCharms showed that an extrinsic reinforcer could actually diminish motivation [50]. His theory was further expanded by Deci and Ryan in the cognitive evaluation theory (CET) that challenged the role of financial incentives in motivation, focusing on the distinction between intrinsic and extrinsic types of motivation [51].

A further refinement of this hypothesis came with self-determination theory (SDT) that distinguished among different types of extrinsic motivation and hence, different effect of financial rewards on different types of extrinsic motivation [19]. According to SDT, the

effect of external (material, tangible not verbal) rewards on work motivation depends on the type of extrinsic motivation, the context and the nature within which such rewards are provided. External rewards are expected to undermine not only intrinsic motivation, but also extrinsic types of autonomous motivation (integrated and identified regulation) when they are perceived as "controllers" of individual's autonomy [52]. However, it appears that the relationship between external tangible rewards and motivation depends vitally on a context in which it is provided. According to the results of the field experiment conducted by Lawler and Hackman, the success of a particular pay plan depended on the process of its introduction, specifically participation by the employees [53]. As Lawler and Hackman [53] explain, when employees participate in development and implementation of such programs they understand them better, are more likely to accept them as being fair, and have stronger commitment to them. Thus, the first key element for designing using financial incentives to increase employee motivation seems to be staff participation.

Theory of goal-setting also provides insights on how to design PBP programs without undermining intrinsic motivation. First of all, PBP program may increase work motivation because it is accompanied by specific indicators or performance goals for the organization or individual. According to the goal-setting theory, people work harder when they have goals than when they do not have them [54]. Secondly, specific and difficult goals were found to be more motivating than general and easy goals [54]. However, use of very specific goals that were attached to bonuses and setting of unattainable goals seemed to actually lower work motivation [54, 55].

SDT [19] also points to possible ways to use rewards without having detrimental effects. According to them, rewards have a positive influence on autonomous motivation if they are provided in organizations where employees experience managerial autonomy support, which is defined as "managers' acknowledging their subordinates' perspectives, providing relevant information in a non-controlling way, offering choice, and encouraging self-initiation rather than pressuring subordinates to behave in specified ways". In laboratory experiments cited by the study [56, 57], financial rewards administered in an autonomy-supportive environment were associated with higher intrinsic motivation and vice versa.

Another element that seems to determine whether financial rewards have detrimental or positive effect on motivation is whether payments are made contingent upon performance or not. Based on the evidence from laboratory studies, it appeared that when payments were made conditional on performance, they reduced intrinsic motivation of the participants, while the same was not true with unconditional payments – they left intrinsic motivation unchanged. Given these

results, Deci came to the conclusion that the effect of financial rewards on intrinsic motivation came not from monetary payments as such but from the way they were administered [58]. It seemed that contingent financial rewards were perceived as controlling, or undermining the personal autonomy.

This seems counterintuitive as pointed out by Locke [59] because this would mean that those employed under a civil service system were more intrinsically motivated than employees of systems based on merit. This also goes against the tenants of principal-agent theory discussed above. However, these results may reconcile findings of studies in developing countries where, on the one hand, low salaries are associated with low motivation and satisfaction among health workers, but on the other hand, there is only weak evidence at best of the positive effect of PBP programs on work motivation and satisfaction.

One last important comment on the potentially negative effects of PBP on health worker motivation in the context of developing countries with often severely underfunded health systems is the difference between "will do" and "can do" factors of motivational processes [47]. As a recent anthropological study points out [43], the PBP programs are based on assumption that the lack of "will do" component of motivation is the primary reason for poor performance. This puts health providers in "double binds" because they often lack resources such as basic infrastructure and medical supplies without which it may be impossible to provide good quality care and achieve the performance targets [43]. This can further undermine already weakened motivation.

In countries such as USA and UK, PBP as a mechanism to improve provider performance has been in existence for quite some time [60]. However, Rosenthal and Frank [60] in their review of PBP programs in USA found little empirical evidence of the impact of these programs on quality of care. According to the authors, this finding of no effect

could have been due to the following reasons: (a) small sample sizes characterizing most of the studies on PBP in USA, (b) the small magnitude of the bonuses relative to the effort required to improve quality, (c) lack of awareness among many participating physicians while the programs were under way, and (d) providing awards only to top few performing medical groups, which made the likelihood of receiving bonuses rather low [60].

A more recent review focusing explicitly on the effectiveness of PBF in improving quality of primary healthcare services conducted by the Cochrane Collaboration concluded that there was insufficient evidence to support or not to support PBF programs [61]. The authors of this review concluded that more rigorous study designs and analytic methods were needed, particularly to account for the selection bias that seemed to be widely present in PBF evaluations. This review included studies from USA, UK and Germany [61]. The 2013 Cochrane review focusing on PBF in low- and middle-income countries, specifically Vietnam, China, Uganda, Rwanda, Tanzania, DRC, Burundi and Philippines, also failed to find strong evidence on the association between PBF and improvements in utilization or quality of health care services [62].

One of the most well-known studies on the success of the PBF in improving use and quality of health care services in a developing country context is the study by Basinga et al on Rwanda [32]. In this study, 166 facilities were randomly assigned to either treatment or control group. To isolate the effect of incentives from the effect of increased resources the authors of the study increased input-based budgets of the control facilities

by the average amount of PBP made to the treatment facilities. It appears that on average, the share of additional funds spent on increased compensation for health workers in both treatment and control groups was very similar: facilities in the treatment group allocated 77% of the PBP funds for personnel compensation, while facilities in the control group allocated 73% of the additional input-based funds to increase staff salaries [32]. Note, that the principles of intra-facility distribution of payments are not described for the control group. The remaining funds facilities spent on facility-level needs such as infrastructure, medical supplies etc. Providers were paid on a quarterly basis. Payments were based on fourteen maternal and child health-care services indicators discounted by the quality of care index.

Basinga et al [32] found statistically significant differences between treatment and control groups in four out of eight impact indicators. According to the conclusions of the study, significant effects were observed for those services for which providers received highest payments or over which they had the greatest control [32]. In short, it seems that there was an increase in services that needed the least efforts from the service provider but provided the highest rewards.

Haiti had a very different study design (see [9] for details) but it also demonstrated positive association between PBP and coverage by key MCH services. Unlike the findings in Rwanda, in Haiti there was a significant and large increase in immunization coverage rates [9]. However, these findings are based on the results of a non-experimental study where NGOs were selected to be in treatment arm because they were

perceived to be ready to graduate into the new payment regime. Moreover, NGOs in treatment arm "aggressively" negotiated for what they perceived to be feasible baselines and targets, while those in the control group tended to accept targets set for them [9]. As Eldridge and Palmer note in their review [41], these achievements may be due to other factors such as technical assistance, capacity strengthening, and overall increase in expenditure by the NGOs. It seems that the effects of PBP were not isolated from the effects of other processes as no data were collected from the control areas [41]. While authors of the study acknowledge a number of reasons why findings from Haiti could be biased, they claim to have used a number of analytical methods to minimize those potential biases and hence, suggest that PBP did have a positive impact on key MCH services, including immunization coverage [9]. However, as Eldridge and Palmer note further in their review, it remains unclear to what extent bonus payments to NGOs reached the actual providers [41].

1.3 Health services in Afghanistan

Public services in Afghanistan have been virtually destroyed by the years of war. Up to 2003, delivery of basic health care services in Afghanistan depended on NGOs with little stewardship from the Ministry of Public Health (MOPH). While NGOs filled an important gap in provision of at least basic health services, health system of the country was fragmented and provision of services was uncoordinated [6]. Most importantly, they were unable to reach large parts of the population: According to the results of the MICS conducted in 2003 [63], skilled birth attendance was at 14.3%, current use of contraception at 10.3%, and DPT-3 coverage rate at 30.1%.

However, since then the Afghan health system has made a strong progress. MOPH has assumed a strong leadership and implemented innovative service delivery models such as performance-based contracting (PBC) accompanied by independent monitoring and evaluation [64]. According to the rules of PBC, in case of poor performance, NGOs may lose their contracts, and MOPH did actually terminate several contracts of underperforming NGOs [6]. At the same time, the Government introduced the Basic Package of Health Services (BPHS) that includes a set of simple cost-effective services delivered mostly at primary care level with a focus on maternal and child health [65]. According to MOPH [66], BPHS is a guaranteed minimum package of services that is to be provided to all citizens regardless of their ability to pay.

These efforts have led to visible improvements in access, financial protection, and utilization of health care services [5, 67]. There has also been some improvement in health outcomes [5]. However, security is still a challenge and is one of the main factors hampering further progress in delivery of health services [65]. Moreover, similar to other developing countries, as described earlier, issues related to health workforce (such as attraction, retention, and performance) further undermine already weak health system and impede progress in health outcomes.

There are three levels of health care services classified in the following way: 1) Primary care provided at the community or village level consists of health posts (HP), sub-health centers (SC), basic health centers (BHC), and mobile health teams (MHT); 2) Secondary

care provided at the district level includes comprehensive health centers (CHC) and district hospitals; and 3) Tertiary care provided at the provincial and national levels consists of national, provincial, and specialty hospitals [66].

At the community level, basic health services are delivered by CHWs from their own homes, which function as community health posts (HP). A HP covers a catchment area of 1,000– 1,500 people, which is equivalent to approximately 100–150 families [66]. Sub-health centers (SC) form a link between HPs and other BPHS levels of service delivery. They were established in remote areas with small pockets of population that could not meet the criteria recommended for BHCs. A SC is intended to cover a population of about 3,000-7,000 with the maximum walking distance of two hours. The SCs provide most of the BPHS services that are available in BHCs including immunization, antenatal care, and family planning. According to the BPHS Guidelines, SCs must have a male nurse and a female community midwife, plus a cleaner/guard [66]. A BHC offers primary outpatient care, including newborn care, antenatal, delivery, and postpartum care, routine immunization, integrated management of childhood illnesses (IMCI), and treatment of malaria and tuberculosis, including DOTS. BHCs are responsible for supervision of HPs in their catchment areas. BHCs cover a population of 15,000–30,000. The minimal staffing requirements for a BHC are a nurse, a community midwife, and two vaccinators. It must have at least one female health worker. CHCs provide a wider range of services as compared to BHCs, including laboratory services and limited inpatient care. In addition to normal deliveries, the CHCs can handle certain complications and provide blood transfusions. They can also manage neonatal infections, complicated cases of

childhood illness and malaria. Each CHC covers a catchment area of about 30,000–60,000 people. A CHC is required to have a male and a female doctor, a male and a female nurse, two midwives, two vaccinators, one CHW supervisor, one laboratory and one pharmacy technician [66].

According to the latest HMIS census [66], there are 59 DH, 395 CHC, 778 BHC, 260 SC, and 10,000 HP. The Basic Package of Health Services is offered at six standard types of health facilities, ranging from community outreach provided by CHWs at health posts to inpatient services at secondary care facilities [66].

1.4 Results-based financing project in Afghanistan

The RBF for Afghanistan has been approved in 2009 and launched in the summer of 2010. Its main goal is "to impact MDGs 4 and 5 by improving coverage of maternal and child health services within the existing health system and without creating unnecessary parallel processes" [10]. Its primary objective is "to increase key maternal and child health outputs, to improve quality of health services and to ensure higher patient satisfaction with health service delivery" [10].

The intervention, which is performance-based payments, was assigned at random where a health facility was the unit of randomization (Figure 1.2). Due to the nature of the intervention there was no blinding of the participants, i.e. facilities and health workers. All facilities have been stratified by type and matched on utilization (average number of outpatient visits per month based on the total number of outpatient visits over the past 12

months) prior to randomization. The evaluation of the intervention was based on two main sources of data: (a) the household survey at baseline and follow-up, and (b) health facility assessment conducted as part of the existing annual national health facility survey described in more detail in the section on data source.

Performance indicators were divided in the following way:

1. Primary care level, plus basic health services at secondary care level provided by CHC: 12 indicators = 9 specific MCH indicators (ANC, SBA, PNC, DPT3, CPR) + TB case detection rate + Equity of care + Quality; and
2. Hospital level: 7 indicators.

RBF scheme included only BPHS facilities and district and provincial hospitals. It did not explicitly include CHWs and the sharing of earned bonuses with CHWs was left at the discretion of NGOs and higher level facilities [68].

The performance payments were allocated in the following way:

1. BPHS facilities (Sub-centers, BHCs, CHCs, outpatient departments of DHs) were to receive $7.7 million (83.7% of the total program funds):
    a. 80% of $7.7 million were to be disbursed at HF level, and
    b. The remaining 20% of the funds were to be disbursed at NGO level and earmarked specifically for the indicator on contraceptive prevalence rate,
2. District and provincial hospitals, the national hospital and Kabul District Hospital were to receive $1.5 million (16.3% of the total program funds).

Unit costs differed by type of services, ranging at BPHS level from $1.5 (per one antenatal care visit to $12 per skilled birth attendance) [68]. However, the original unit costs were increased one year after the start of the program and put into effect in the last quarter of 2011 [69]. Each facility in the treatment group was likely to earn some amount of bonus payments because improved performance was defined according to each facility's own baseline. According to the MoPH [70], regardless of facility performance relative to others if it did not exceed its own baseline value, it would not receive the bonus. For example, if facility A in Quarter 1 reported 10 skilled birth attendance cases and in Quarter 2 it reached 20, while facility B in Quarter 1 had 20 skilled birth attendance cases and in Quarter 2 it stayed at that level, facility A would receive bonus payments while facility B would not. Also, the payment would be "discounted" by the quality of care (0 - 1) as measured by a quarterly score on the national monitoring checklist (NMC) that focused on structural indicators of quality and was jointly carried out by supervisors from the implementing organization and members of the provincial health office.

An important element of the program design that was likely to influence motivation and, consequently, performance, was the distribution of incentives within facilities. The distribution of incentives was left at the discretion of health facility staff through a consultative process with NGO field managers [71, 72]. This meant the implementing NGO would sign a contract with the health facility about the principles of distribution of performance incentives within the facility. There were three major mechanisms on the distribution of performance incentives at facility level:

- Amount of individual incentive proportional to salary amount;
- Amount of individual incentive distributed equally regardless of staff position or performance; and
- Amount of individual incentive based on contribution to performance indicators.

However, the NGOs were free to come up with some other mechanisms if the three described above were not acceptable to them.

Most of the indicators and payments against them were to be made on a quarterly basis. Routine health information collected and aggregated on a quarterly basis was used for performance monitoring. NMC was used to monitor quality of care, which was also collected on a quarterly basis. The equity and Balanced Scorecard (BSC) indicators (for more details on BSC see [40]), however, were collected on an annual basis through the existing health facility assessment. For the quarterly indicators submitted through HMIS, there was a two-step system of verification. First, there was verification of data submitted to HMIS against facility records. Second, verification of facility records was done through household interviews by community-based organizations. The process of collection and verification of data from health facilities and up resulted in significant delays of payments, particularly at the start of the program.

1.5 Baseline

As described in the section above, this is a cluster randomized trial. The unit of randomization was a health facility. In accordance with principles of randomization, it is assumed that treatment and control groups are equivalent with respect to all measured

and unmeasured variables, except the treatment itself [73]. This assumption is likely to be violated in small sample studies, and baseline study can be used to confirm whether randomization process was a success or not. However, given the size of the current trial (442 clusters in 11 provinces), randomization process was expected to produce equivalent groups [73, 74].

The methods and results of the baseline survey, described in detail in the "Impact Evaluation of the Results-Based Financing Intervention in Afghanistan" [75], are summarized below. The data was collected between 2009 and 2010 as part of the NHSPA. Thus, the overall sampling strategy and sample size have been determined outside of the study, i.e. by the parent study. A primary sampling unit was a health facility. The baseline study sample was 53 matched pairs, or 106 facilities. This included 362 health workers and 1,013 patients.

The study was based on intention-to-treat analysis at cluster level [76]. The test of normality was conducted on cluster-level means. The intervention effect was estimated using difference in means, and 95% CI for the difference in means were calculated on an original scale for variables with normal distribution and on logarithmic scale for variables with skewed distribution. Paired t-test and the signed rank test of cluster means at 5% significance level were used to test the intervention effect.

The indicators measured at baseline were part of the national health sector Balanced scorecards [40]. The findings confirm that the randomization produced two comparable

groups (Table 1.2). No statistically significant differences were found among the 18 indicators in four domains: (a) Patients and community, (b) Staff, (c) Physical capacity, and (d) Service provision.

1.6 Thesis research objectives and conceptual framework

The goal of this dissertation research is to examine the effect of performance-based payments on health worker motivation and quality of care in Afghanistan. The research objectives are as follows:

1. To develop a multidimensional work motivation scale and assess its psychometric properties in Afghanistan (Paper 1);
2. To assess motivation among health workers in facilities assigned to performance-based payments program as compared to health workers in facilities assigned to control in Afghanistan (Paper 2); and
3. To assess quality of care for patients observed in health facilities allocated to performance-based payments program as compared to those observed in health facilities allocated to control in Afghanistan (Paper 3).

The conceptual framework for this dissertation is presented in Figure 1.3, at the end of this chapter. The framework is adopted from Franco et al [47] study of the health worker motivation in Jordan and Georgia. It is drawn using structural equation modeling notation to distinguish latent variables such as motivation drawn as circles from observed variables drawn as squares. As Franco et al [47] note, there are broader factors (outside the organization and individual) that impact worker motivation. Organizations do not exist in vacuum and are influenced by broader social, political and economic factors. For

example, in Afghanistan religion plays a very strong role in society and influences organizational culture and individual motivation. While the framework presented by Franco et al [47] shows only direct arrow from social factors to motivation, it seems more reasonable to also assume that these broader contextual factors directly influence the individual performance as well motivation. For example, widespread insecurity in Afghanistan is likely to have a direct effect on performance as well as an indirect effect through motivation.

Organizational factors that define the work environment include structures, resource availability and culture. Structural factors include formal scope of work for each employee and department, promotion rules, procedures for hiring and firing and other rules that constitute the formal system of task and authority [24]. Cultural factors include values and norms of the organization [24]. For example, in a setting such as Afghanistan, a formally flat organization can be very hierarchical in reality due to cultural norms. There is a double headed arrow between organizational factors and performance-based payments. As described in earlier sections, in some countries, such as Rwanda, part of the PBP earned by the facilities were allocated for facility-wide investments such as infrastructure, equipment and medical supplies, thus influencing resource availability in an organization (arrow from PBP to organizational factors). In Afghanistan the entire amount earned by the facility was divided among health workers; however, conceptually this link exists. The arrow from organizational factors to PBP is drawn to show that organizational structural and cultural factors influence the way PBP are distributed (intra-facility distribution) and whether an individual health worker actually receives such

payments. As discussed in the section on RBF in Afghanistan, different organizations there chose to distribute PBP differently (equal amount to every staff member, in proportion to existing salary, or only to those staff members who contributed directly to performance indicators). This decision is a result of a complex set of organizational factors.

Individual worker characteristics such as gender, health worker type (doctor, nurse etc) or level of training, and individual capability (knowledge and skills) are important determinants of motivation and performance at the individual level [37, 47]. Motivation influences health worker performance and two otherwise similar doctors will perform differently, depending on their level and type of motivation [77]. Worker capability also has direct influence on worker performance, not just through motivation. There are different consequences of worker performance, depending on the type of financing system and organizational structure. In systems where workers are paid monthly salaries based on their level of training and years of experience, individual performance is unlikely to have significant consequences. There is a dashed arrow back from the consequences of worker behavior to worker motivation demonstrating the link between motivation in period two and consequences of one's performance in period one. The dashed arrow is used to show that the study of this feedback mechanism requires longitudinal data, which is not available in this study.

1.7 Relationship between thesis papers and the project from which the data derive

The data used in the three papers that comprise this thesis derive from data collected as part of the Monitoring and Evaluation Technical Assistance for Strengthening Health Activities for the Rural Poor (METASHARP) Project implemented by the Johns Hopkins Bloomberg School of Public Health (JHSPH, referred to as JHU in Kabul) in collaboration with the Indian Institute of Health Management Research (IIHMR). The project began in 2009 and closed in 2013. The core activity of the project was the National Health System Performance Assessment (NHSPA) that focused on the development and use of the Balanced Scorecards based on the annual health facility surveys. In addition, JHSPH and IIHMR consortium provided technical assistance to MOPH in conducting and analyzing national household surveys. The consortium also built technical capacity of the Ministry staff to analyze and interpret health data. Moreover, as part of the METASHARP, JHSPH and IIHMR were required to conduct community verification activities of the HMIS data and the evaluation of the RBF project based on health facility and household surveys at baseline and follow-up.

I started working on the project in the spring of 2009 when I contributed to the analysis of the National Risk and Vulnerability Assessment (NRVA) household survey. In June 2010, I joined the team in Kabul and took the lead in designing the household survey instrument and training materials for the RBF baseline evaluation. In 2011, following the data collection, I led the analysis of the RBF baseline household survey. I also contributed to the revision of the Balanced Scorecard instruments for the NHSPA,

focusing on the health worker questionnaire (form F5). In February 2012, I returned to Kabul as a core member of the field team to assist with data collection and analysis of the NHSPA 2012 – 2013 round, a main data source for this study. Upon return from Kabul, I continued my involvement with the project and assisted in the analysis of the NHSPA data, contributing to the write-up of the national Balanced Scorecard Report 2012 – 2013 as well as the final RBF evaluation report.

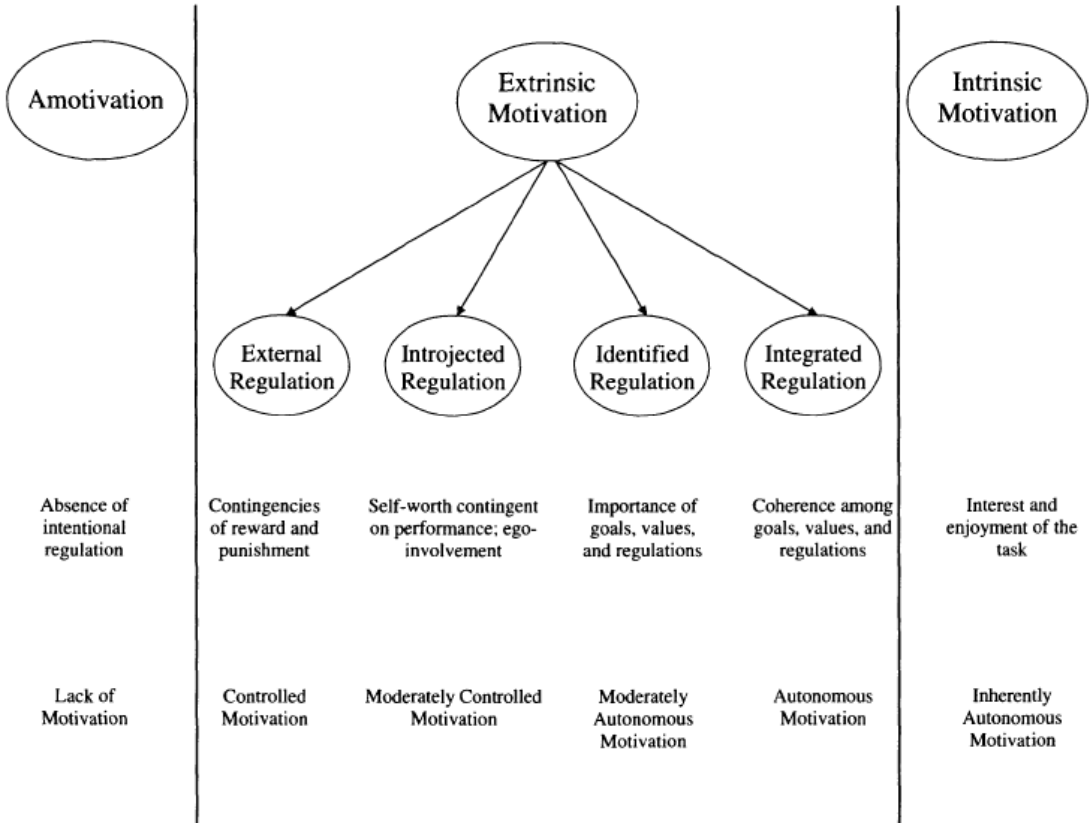
1.8 Organization of this document

The rest of this thesis is organized as follows:

- Chapter 2 describes the data sources used, including sampling and data collections methods
- Chapter 3 assesses the reliability and validity of the newly developed work motivation scale in Afghanistan (Paper 1)
- Chapter 4 examines the effect of performance-based payments on health worker motivation in Afghanistan (Paper 2)
- Chapter 5 examines the impact of performance-based payments on quality of care in Afghanistan (Paper 3)
- Chapter 6 summarizes the findings and draws conclusions and recommendations

The main tables and figures for each chapter appear at the end of the chapter. Additional tables appear in appendices.

Figure 1.1 Structure of motivation according to the Self-Determination Theory



Source: Gagne and Deci, 2005

Figure 1.2 Study design of the Results-Based Financing Project in Afghanistan, BPHS facilities

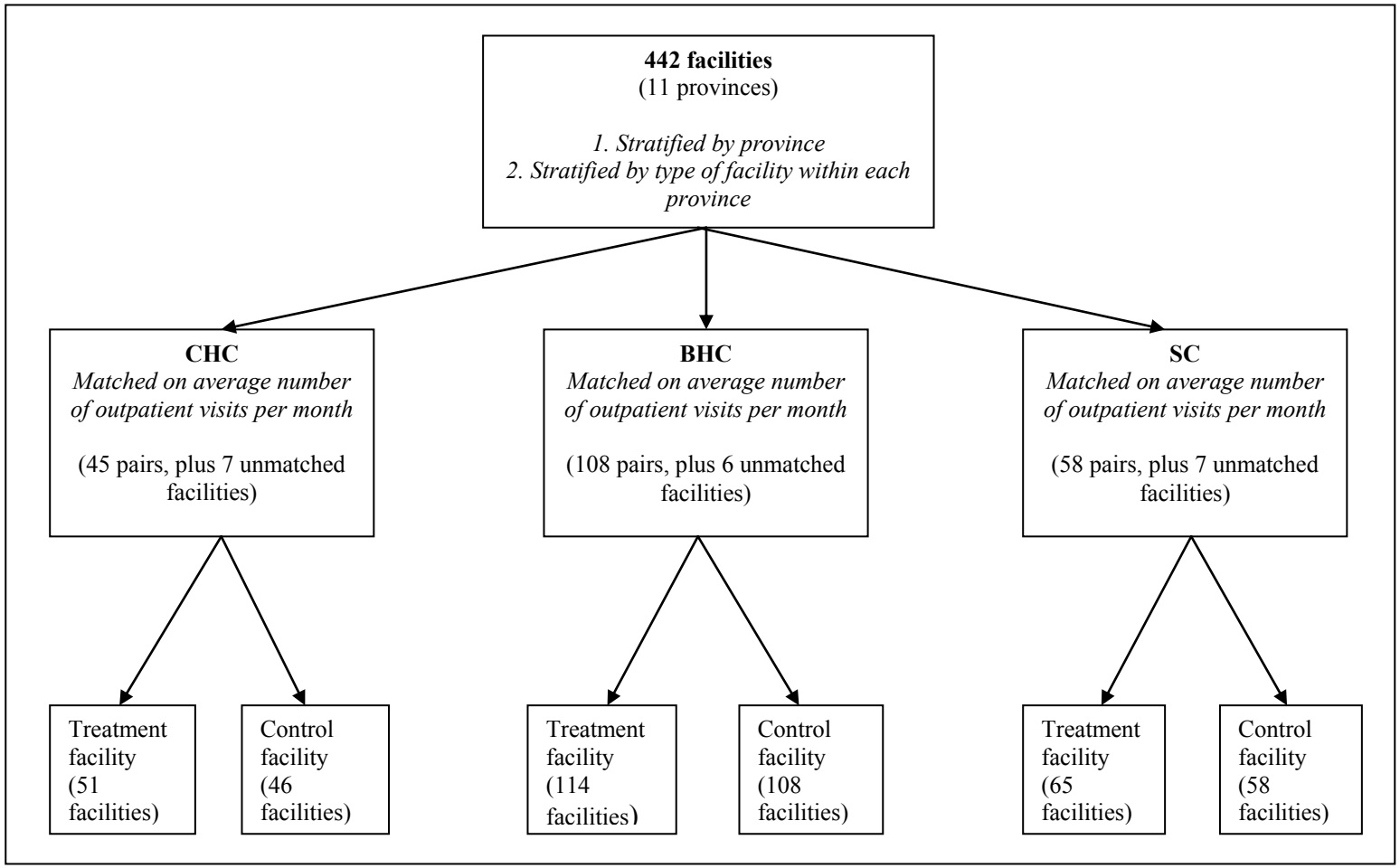
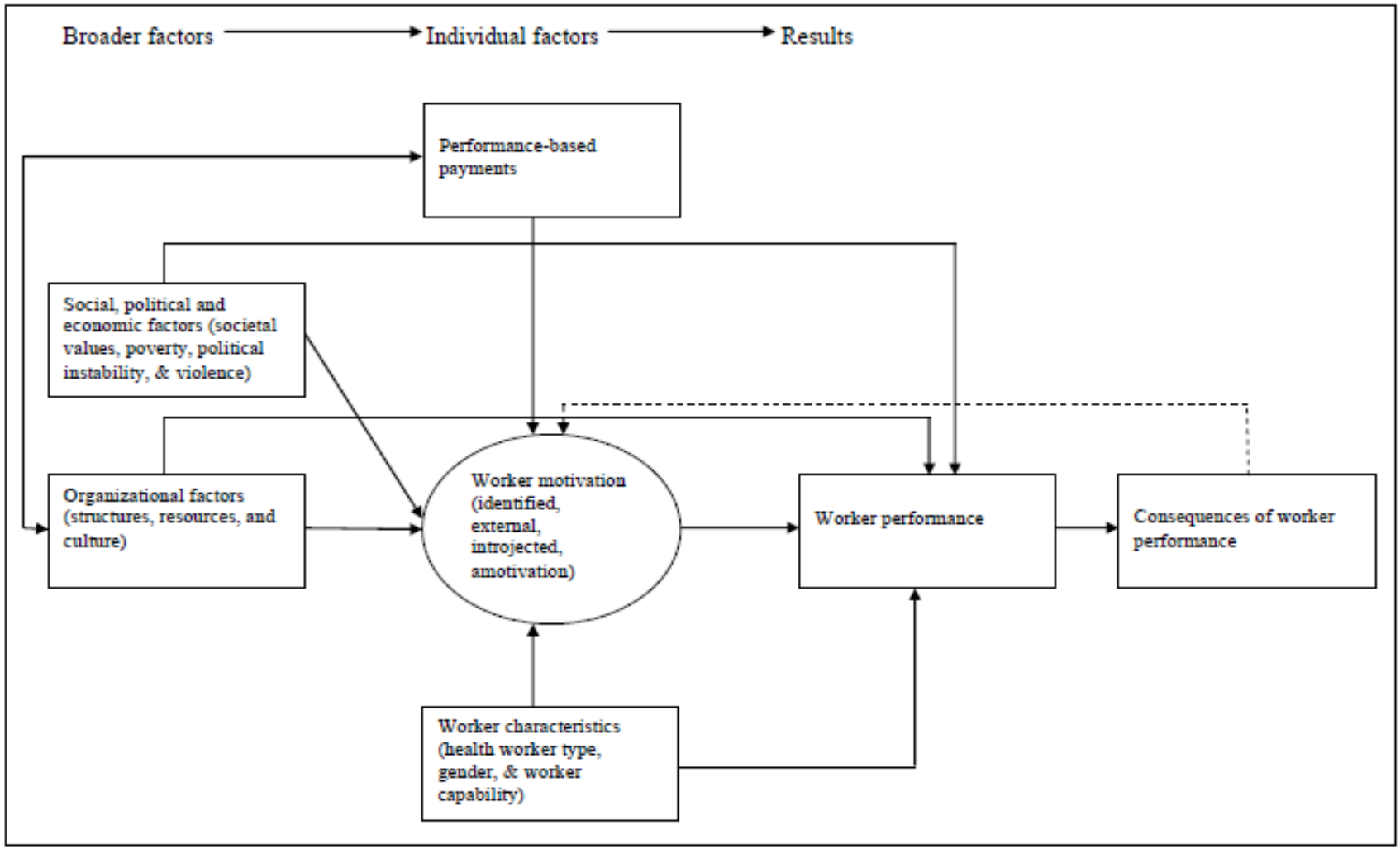


Figure 1.3 Conceptual framework



Source: Adopted from Franco et al, 2000

Table 1.1 Overview of motivation theories

Content theories	
The Need Hierarchy	Maslow, 1943
ERG Theory: Existence, Relatedness, Growth	Alderfer, 1972
Two-Factor Theory: Motivators vs. Hygiene Factors	Hertzberg, 1959
Learned Need Theory: Need for Achievement, Need for Power, and Need for Affiliation	McClelland, 1976
Process theories	
Equity Theory	Adams, 1965
Expectancy Theory: Job Outcomes, Valences, Instrumentality, and Expectancy	Vroom, 1964
Reinforcement Theory or Operant Conditioning: Stimulus, Response and Consequence	Skinner, 1969
Cognitive Evaluation Theory: Intrinsic and Extrinsic Motivation	Porter and Lawler, 1968; deCharms, 1968
Goal Setting Theory	Locke and Latham, 1984
Social Cognitive Theory (self-efficacy)	Bandura, 1977
Self-Determination Theory	Deci and Ryan, 2002

Source: Shortell and Kaluzny, 2006

Table 1.2 Baseline characteristics, cluster-level analysis

BPHS Balanced Scorecard		Treatment Arm		Control Arm		
Domain A	**Patients and Community**	**Mean**	**95% CI**	**Mean**	**95% CI**	**p-value**
1	Overall Patient Satisfaction	76.4	73.53; 79.32	77.8	75.19; 80.49	0.42
2	Patient Perceived Quality of Care Index	76.2	74.01; 78.37	76.9	74.65; 79.20	0.60
3	Community Involvement and Decision Making Index	98.1	95.46; 100.76	92.5	86.17; 98.73	0.15
Domain B	**Staff**					
4	Health Worker Satisfaction Index	70.8	68.87; 72.77	69.9	67.71; 72.08	0.41
5	Salary Payment Current	58.8	46.95; 70.66	63.7	53.39; 73.97	0.30
Domain C	**Physical Capacity**					
6	Equipment Functionality Index	87.7	83.91; 91.41	85.4	82.09; 88.76	0.24
7	Drug Availability Index	84.9	78.43; 91.38	84.9	79.40; 90.41	0.57
8	Family Planning Methods Availability Index	94.3	90.60; 98.08	96.2	91.87; 100.58	0.27
9	Laboratory Functionality Index	69.7	57.80; 81.59	61.4	46.56; 76.17	0.27
10	Minimum Staffing Index	81.1	70.24; 92.02	77.4	65.71; 89.00	0.56
11	Staff Received Training (in the past 12 months)	14.8	10.94; 18.65	12.1	9.00; 15.28	0.35
12	HMIS Index	78.6	71.85; 85.38	78.0	71.50; 84.47	0.56
13	Clinical Guidelines Index	78.9	73.13; 84.58	83.0	77.73; 88.31	0.20
14	Infrastructure Index	59.7	52.80; 66.59	60.2	53.38; 67.01	0.75
15	Patient Record Index	63.1	59.41; 66.77	63.9	61.40; 66.33	0.51
Domain D	**Service Provision**					
16	Patient History and Physical Exam Index	68.3	65.08; 71.51	67.8	64.60; 70.93	0.78
17	Patient Counselling Index	30.1	26.25; 33.97	32.8	29.40; 36.27	0.21
18	Time Spent with Client	12.9	8.50; 17.37	8.3	4.81; 11.83	0.10

Note: The overall sample size is 53 pairs. P-values are calculated with Wilcoxon matched pairs signed ranks test for non-normally distributed continuous or binary outcomes and with paired t-test for normal continuous outcomes. Confidence interval estimates are calculated with paired t-test. Laboratory functionality index applies to CHCs only; the sample size for this indicator is 12 pairs.
Source: JHSPH/IIHMR, 2013

2 Chapter 2: Data sources used

The data for this study was derived from the annual National Health Services Performance Assessment (NHSPA) health facility survey implemented in Afghanistan over the period of nine months between 2012 and 2013 [40]. The survey was based on a multi-stage stratified probability sampling approach where primary sampling unit was a health facility:

- Stratification by province,
- Stratification by type of health facility within each province,
- Selection of facilities,
- Selection of individual health workers, and
- Selection of patients.

Individual health workers and patients were secondary sampling units. Within each facility four health workers classified as clinical staff (doctors, nurses, assistant doctors, midwives, community midwives, and vaccinators) present at the day of the survey were randomly selected for the survey. If fewer than four clinical staff were present at the facility, then other facility staff members, including community health supervisors, pharmacists, cleaners and others, were interviewed for the survey. However, all non-clinical workers with the exception of community health supervisors were excluded from the analysis for this thesis.

Patients were selected independently of health workers. Within each sampled facility, patients were stratified by age group (less than five years of age and five years of age or older) and selected based on a systematic random sampling approach as described below:

- If less than 10 new patients were expected in a day in each age stratum, then each eligible patient was selected until 5 observations of consultations involving patients in that age stratum were completed;

- If 11 to 15 new patients were expected in a day in each age stratum, then every second eligible patient was selected;
- If 16 to 20 new patients were expected in a day in each age stratum, then every third eligible patient was selected;
- If more than 20 new patients were expected in a day in each age stratum, then every fourth eligible patient was selected.

According to NHSPA sampling methodology, at provincial level there was equal allocation with twenty five facilities selected per each province. Within each province, allocation was in proportion to stratum size (where stratum is the facility type) with fifteen BHCs, five CHCs and five SCs [40]. This sampling methodology was followed in 2009 – 2010 for the baseline study and resulted in 213 facilities across eleven RBF pilot provinces. Of these, only 106 facilities were selected together with their matched pairs (resulting in 53 pairs) as the NHSPA survey had to start before the RBF matching and randomization were completed.

As the RBF evaluation was done within NHSPA, the 2012 – 2013 survey was also based on the above sampling methodology. Thus, following the allocation procedure per stratum described in the previous paragraph, 25 facilities per province were randomly selected. Once these were selected, 106 facilities which were randomly selected at baseline in 2009 – 2010 were purposefully added to the sample for the follow-up survey, i.e. 2012 – 2013. This was done to ensure that same 53 pairs surveyed at baseline were also surveyed at follow-up, as required by the project evaluation design. This addition impacted four out of eleven pilot provinces. These are the following: Parwan (36 facilities), Takhar (26 facilities), Kunduz (31 facilities), and Saripul (27 facilities). The thesis is based on 2012 – 2013 data and thus includes all 256 facilities, i.e. facilities randomly selected for the 2012 – 2013 NHSPA round and those selected randomly in 2009 – 2010 but purposefully added in 2012 – 2013. Also, it should be noted that in Kandahar, due to security concerns only eight facilities were selected for RBF pilot. Thus, while under NHSPA 25 facilities were surveyed in Kandahar, only eight of these were included in this study. Facility weights were used when estimating means and proportions to adjust for the unequal probability of selection, resulting from this sampling methodology.

The NHSPA survey of BPHS facilities consists of seven instruments, plus a separate instrument for health posts. F1 – F2 are instruments used for patient observations. F3 – F4 are used for exit interviews with patients or their caretakers. F5 is the health worker questionnaire. F6 is a special short questionnaire for community health workers (CHWs). F7 is designed to assess facility level characteristics, such as infrastructure, availability of

medical supplies, and staffing levels. Data used for Papers #1 and #2 are derived from forms F5 and F7. Data used for Paper #3 are derived from forms F1, F2, F3, F4, F5, and F7. Thus, for Papers #1 and #2, facility data from F7 were matched with health worker data from F5. This resulted in a sample of 256 health facilities with 805 health workers, or 3.1 health workers per facility.

For Paper #3, the data set created for Paper #2 was further matched with the data derived from F1 – F4 forms. This resulted in a sample of 233 health facilities with 255 health workers, and 2,180 patients. On average, there were 1.1 health workers per facility, 8.5 patients per health worker, and 9.4 patients per facility. Given that not all those health workers that were selected for health worker questionnaires were also observed, and not all those health workers who were observed were selected for health worker questionnaires, the sample of health workers for Paper #3 (n=255) was substantially smaller than the sample for the study in Paper #2 (n=805).

More than 90% of the data was collected by survey teams that have been conducting surveys for the NHSPA project for several years. Moreover, all members of these teams had medical degrees. A small portion (6%) of data was collected through community survey teams due to security considerations. These were almost exclusively in non-RBF pilot provinces. The community survey teams received a short training but they did not have medical degrees and were hired only for this survey.

3 Chapter 3: Reliability and validity of the newly developed work motivation scale among health workers in Afghanistan (Paper 1)

Abstract

Background: Access to health services depends to a large extent on availability of the health workforce. Shortage and low performance of individual health workers are often linked to low motivation. At the same time, it seems that intrinsically motivated health workers can make a positive difference in access to and quality of health services, even in settings with few resources. However, there has been some difficulty in producing strong evidence on the impact of motivation on health workers' performance. To a large extent this is due to measurement challenges as latent constructs such as motivation cannot be directly observed or measured.

Method: Based on the Self-Determination Theory of motivation, a four-subscale 15-item scale was developed and tested as part of the National Health Services Performance Assessment in Afghanistan. The sample (n=2,413), including speakers of two main languages of Afghanistan, Dari and Pashto, was split into four parts. Exploratory factor analysis was performed on sample one, which contained only Dari speakers (n=431). Based on these results, confirmatory factor analysis was performed on sample two (n=439), which also contained only Dari speakers. This model was cross-validated using CFA on a third sample of Dari speakers (n=805). Finally, the generalizability of the scale was tested on Pashto speakers (n=738). Reliability of the scale was assessed using ordinal alpha based on polychoric correlation matrix. Construct validity of the scale was assessed using the following variables: intention to quit, performance assessment and feedback,

staff voice and equity of rewards, and general motivation. Standard error computations used a sandwich estimator to account for non-independence of observations as a result of cluster sampling.

Results: The scale demonstrated good psychometric properties. Three out of four subscales had ordinal alpha of 0.70 or higher. The results of the multiple group analysis demonstrating the difference in factor means for amotivation between respondents intending to quit and those intending to stay provided evidence for construct validity of the scale. The factor mean for amotivation was significantly lower (p-value<0.05) in the group that was intending to stay, as compared to the group that was intending to quit. Staff voice had statistically significant (p-value<0.05) and positive association with identified, external and introjected regulation factors. Equity of rewards appeared to be a strong predictor of external regulation and introjection (p-value<0.05). Amotivation had a negative correlation (-0.30) with the question on general motivation, providing evidence of discriminant validity of the scale.

Conclusion: The work motivation scale developed in this study demonstrated good psychometric properties. The scale developed using the Dari language showed generalizability to Pashto language as well. Availability of a short (15-item) scale validated in non-Western context using a large representative sample is an important step in improving our understanding of health worker motivation and designing more nuanced human resource policies in developing countries.

3.1 Introduction

Access to health services depends to a large extent on availability of health workforce. There is strong evidence that links staffing levels to both service delivery and health outcomes [78]. As the WHO paper on "Priorities and Strategies in Human Resources for Health" has put it, "The single most important determinant of the performance of health services delivery systems is the performance of health workers" [79]. A wide range factors (see Figure 1.3 Conceptual framework), such as the overall security situation in the country, infrastructure of the facility, its organizational culture, training of individual health workers and their motivation, have been identified as determinants of performance [36, 39, 77, 80]. On the one hand, there is some evidence suggesting that low motivation contributes to shortage and low performance of individual health workers, weakening the entire health system [45, 81]. On the other hand, it seems that intrinsically motivated health workers can make a positive difference in access to and quality of health services even in settings with hard resource-constraints [77, 82].

However, in contrast to more easily measurable factors such as training or infrastructure, there has been some difficulty in producing strong evidence on the impact of motivation on health workers' performance. To a large extent this is due to challenges posed by constructs such as motivation, when it comes to measuring them [83]. While errors in measurement are widely present even in physical attributes such as weight and length, they can be measured directly and their measures can be replicated to provide confidence in their accuracy. Motivation is a latent construct, and this implies that it is not directly observed, it can only be inferred from observable behaviors or attributes [84].

The primary objective of this study was to develop a multidimensional work motivation scale and assess its psychometric properties in Afghanistan. The immediate application of the instrument is for the assessment of the impact of a performance-based financing project on health worker motivation in Afghanistan (see Chapter 4). Based on the review of the literature on motivation, as described in Chapter 1, it was hypothesized that work motivation is a construct that consists of five factors, indicating the degree of internalization of goals and values of an organization. The hypothesized factors were: integrated regulation, identified regulation, introjected regulation, external regulation, and amotivation (see Table 3.1 for more explanation of these terms). These factors reflect the degree to which the intent to act is autonomous (factors 1 and 2) controlled (factors 2 and 3), or lacking completely (factor 5).

While theoretical understanding and empirical measurements of motivation in general and work motivation in particular have been growing, they have been largely limited to North American and Western European settings. In developing countries, there is a scarcity of rigorous quantitative studies on work motivation based on clear theoretical framework [48, 85]. Not rarely, studies, while distinguishing constructs such as "motivation" and "satisfaction" in their theoretical review, appear to combine them in their measurement scales where some items seem to ask about motivation ("Only do this job to get paid") and other items ask about satisfaction ("I am satisfied that I accomplish something worthwhile in this job") [83, 86]. They are also often limited by one or more methodological issues, including small sample size, non-probability based sampling, low

response rate, or statistical methods of analysis that do not take the full advantage of new developments in psychometrics [83, 86-89]. A systematic review of motivation in developing countries by Willis-Shattuck et al [81] uncovered only eighteen studies that focused on health worker motivation and intervention(s) linked to it that met their quality criteria, including use of primary data. Many excellent studies are based on qualitative methods [45, 46, 90-92], which are necessary in developing initial understanding of motivation and related to it constructs from the perspective of health workers in a given setting, but have limited generalizability and cannot test hypotheses with large representative samples [90, 93].

Studies focusing on motivation and satisfaction of health workers in Afghanistan or Central Asia in general are almost non-existent. Using key words such as "work motivation" and "scales", PsycINFO – a primary database for obtaining access to scales and studies on their application – was searched. No language limitations were imposed. The only limitation applied to the search was "published between 2000 and 2013". This resulted in forty-seven publications. None of these had Afghanistan or any other Central Asian country in either title or abstract. There were two publications on Pakistan, one on teachers and one on banking sector employees, and three publications on India, published in Indian journals. Searches using terms such as "motivation", "satisfaction", "work motivation" in PubMed also did not result in studies on work motivation in Afghanistan or other Central Asian country, although there were a few studies on Pakistan and India [16, 91, 94]. While this quick search was limited to only two databases, it indicates the scarcity of studies on work motivation in these countries. A systematic review of

motivation in developing countries by Willis-Shattuck et al [81] also points to a shortage of rigorous studies of motivation in this region. Out of eighteen studies included in the review, only one was from Central Asia or Afghanistan. At the same time, large resources are currently being spent in this region through projects such as Results-Based Financing in health (RBF), which aim at improving health worker motivation through financial incentives (see Chapter 1 for more details on the project in Afghanistan) [95]. Thus, improving our understanding of motivation and developing more rigorous methods to measure it are important tasks for researchers and practitioners working on improving health in this region.

3.2 Methods

3.2.1 Data source

The data for this study was derived from the annual National Health Services Performance Assessment (NHSPA) health facility survey implemented in Afghanistan over the period of nine months between 2012 and 2013 [40]. The survey was based on a multi-stage stratified probability sampling approach where a primary sampling unit was a health facility (Figure 3.1):

- Stratification by province,
- Stratification by type of health facility within each province,
- Selection of facilities,
- Selection of individual health workers.

Individual health workers were secondary sampling units. Within each facility four health workers classified as clinical staff (doctors, nurses, assistant doctors, midwives, community midwives, and vaccinators) present at the day of the survey were randomly selected for the survey. If fewer than four clinical staff were present at the facility, then other facility staff members, including community health supervisors, pharmacists, cleaners and others, were interviewed for the survey. However, all non-clinical workers with the exception of community health supervisors were excluded from this analysis. This resulted in a sample of 775 health facilities with 2,413 health workers, or 3.1 health workers per facility. The unit of analysis is a health worker.

3.3 Measures / Variable construction

Work motivation scale

The main steps in the construction of the scale are described in Figure 3.2. Using existing literature on motivation, the construct of motivation was defined. Self-Determination Theory of motivation (SDT) formed the theoretical framework for the proposed instrument. Motivation was hypothesized as a multi-dimensional construct with five distinct factors, each forming a separate subscale (Table 3.1).

Content and face validity

Having defined the construct, a search was done for reviews and applications of existing scales on motivation, focusing on studies conducted in similar social and economic context. At this stage, given that the terms “satisfaction” and “motivation” have been used interchangeably in many of the studies, satisfaction scales or studies were included

as well. Next, main domains from different instruments were pooled together and mapped to identify duplications, common themes, and new areas. Following this, a very broad pool of items was collected and then grouped under the identified domains. Many of the items were very similar and hence, only one of them was retained. As a result, a pool of items was developed from other measures that met the minimum criterion of face validity given the adopted theoretical framework.

These items were then grouped into five categories reflecting the hypothesized dimensions of motivation as a construct. Several items, particularly for external regulation subscale, had to be introduced for the first time to reflect the existing external rewards structure in Afghanistan. In consultation with an Afghan advisor, the author conducted close review of the items in the initial pool. It was found that (a) many of the items from different instruments were very similar in content and thus, using one of them would be sufficient; (b) several items were worded in ways that did not seem appropriate for the given context. For example, the Work Extrinsic and Intrinsic Motivation Scale (WEIMS) asked a question "Why do you do your work?" and offered a series of responses, including items such as "Because I want to be a "winner" in life" and "I ask myself this question, I don't seem to be able to manage the important tasks related to this work" [23]. First item did not seem appropriate because calling oneself a "winner" and stating that as one's goal for choosing a certain type of work seemed to be a rather foreign idea in Afghanistan. Second item had a very complex structure. Based on this review, 31 items were selected and shared with a larger group of in-country and JHU

experts. Finally, 20 items were selected for translation and pretesting of the scale (see Appendix A, Table 7.1).

This preliminary scale was then translated into the local languages. Following this, the instrument was tested with 50 health workers as part of a larger field-testing exercise for the NHSPA instruments in a hospital in Kabul. The items were checked for missing values and correct direction of responses. The scale was then translated back into English, and some items were reworded to achieve simpler sentences. This scale was then piloted as part of the NHSPA 2011. In this preliminary scale (see Appendix A, Table 7.1), items 251, 253, 261, 266, and 267 were adopted from the motivation scale developed and applied among Kenyan health workers [86]; items 254, 255, 265, and 269 were derived from the original Job Diagnostic Survey [96]; and item 258 was taken from the Job Satisfaction Survey [97]. The remaining 10 items, mainly on self-worth contingencies and external rewards, were newly introduced, although several of them were based on items such as those tested among physicians in Pakistan [91]. Item 268 was completely new and specifically introduced to measure the importance of performance-based payments (PBP) in externally regulated motivation, because as described in the introduction, PBP were being introduced in Afghanistan to improve health worker motivation.

The respondent was provided with a statement that referred to their current job and she was asked to indicate the level of her agreement with that statement. The response scale was a four-point Likert scale that had been well tested in the context of Afghanistan [35].

To maximize the readability of the scale, items were assessed using the Flesch reading ease formula [98]. This was done using Microsoft Office Word Program that provides readability statistics as part of the spelling and grammar check. The Flesch reading ease score for all of the items taken together was 71.7 and the Flesch-Kincaid grade level was estimated as being equal to the 7th grade. In addition, given the literacy levels in the population the reading and comprehension of the tool was tested with actual health workers as described below.

In parallel, two focus group discussions with seven health workers in each from two different facilities in Kabul were conducted. The guide for the focus group discussions is attached (Table 7.2). These discussions showed that all health workers understood work motivation as something that causes a worker to act in a certain way. However, for the majority it was something that was externally determined and was often equated with encouragement, financial rewards, promotion and support. This is very similar to findings of the qualitative study in Kenya and Benin [45], according to which over 50% of health workers in Benin equated motivation with prospective "encouragement" with one fourth of these explicitly mentioning financial encouragement, while another 40% considered "being motivated" as having the necessary means to work and get recognition. As the authors of this study put it, the majority understood motivation as “an incentive, and not as a state of mind” [45]. This is very similar to results of the focus group discussions with Afghan health workers. Moreover, it seemed that the word “tashweeq”[3] that was

[3] Tashweeq means "evocation of shauq", shauq being a word for desire, zeal, inclination etc. Especially combined with the verb "kardan" (to do), tashweeq kardan, it means to encourage. Combined with the verb "shodan" (to become), tashweeq shodan, it might acquire an intrinsic aspect. However, in Persian, the word "angize" means motivation.

frequently used in these discussions and seemed to be used in daily language in Dari was the word for encouragement, while a more accurate word for motivation, “angize”, was used more rarely, although it seemed to be understood by all health workers [99].

At the same time, there were several participants who viewed motivation as “an independent desire to perform” that came from understanding consequences of one’s actions and desire to do something meaningful. Religion proved to be one of the important motivating factors, along with the desire to help people or family, respect in the community, verbal appreciation from superiors, the fairness and competence of the superiors, and financial rewards. Interestingly, financial rewards were connected to religion: Several participants said that being able to earn “Halal” or lawful money was one of the motivating factors. Thus, these discussions show that motivation is viewed as both intrinsically and extrinsically determined, although it seems that extrinsic factors are more dominant.

In addition to the focus group discussions, the questionnaire was tested in a small group of 7 different types of health professionals through the following steps:

1. The interviewer read out each question and asked them to repeat it back. This was recorded in the column next to the corresponding question.
2. Each health worker filled out a clean questionnaire form.
3. The interviewer then reviewed the answers and asked the respondent to explain their answers. These were recorded next to the corresponding question.

Overall, it appeared that items were well understood by the respondents. However, three out of seven respondents seem to have understood item 257 on opportunities for promotion as opportunities for improvement or training. Having reviewed the back translation, it became apparent that "promotion" had been translated as "improvement" and thus, while some respondents understood it correctly, others did not. Item 265 ("It is hard for me to care very much about whether or not the work gets done right"), which was intended to measure amotivation, seemed to have been misunderstood by all of the respondents. This item was subsequently dropped.

Pilot study

As described above the 20-item scale was piloted as part of the NHSPA 2011 (Figure 3.2). Using the data from 11 RBF pilot provinces (n=811), direction and response distribution of each item of the scale was examined. After examining overall distribution, exploratory factor analysis accounting for clustering through robust standard errors with GEOMIN oblique rotation was used to uncover the underlying dimensions of the scale. Given that it was clustered ordinal data, WLSMV was used as an estimator. Based on the magnitude of eigenvalues, it seemed that four factors should be retained (eigenvalue>1). Together, they accounted for 52.3% of the total variance. Based on the chi-square test of model fit the null hypothesis of a single-factor was rejected. However, the chi-square test is sensitive to sample size and can lead to rejection of a model that may be only trivially misspecified. Thus, other global fit indices were examined. Based on the global fit indices, it appeared that there was no significant improvement in the model after a two-factor model was introduced. However, based on factor loadings and residual variances, a

four-factor model was finally selected. The internal consistency of the subscales was assessed using Cronbach's alpha for ordinal data [100]. While the factor on identified regulation performed well (alpha coefficient of 0.79) and the factor on external regulation was satisfactory (alpha coefficient of 0.68), the factor on amotivation was below the desired standard (alpha coefficient of 0.57) and the factor on introjected regulation with only two items required additional items [101].

This scale was then reviewed by an external Dari to English translator with a graduate education in psychology [102]. Based on his review, findings of the focus group discussions, and the results of the EFA, it was determined that questions not performing as expected would be reworded for clarity and several new items would be introduced into the scale. Also, items originally designed to measure integrated and identified regulation factors were combined into one factor based on the results of the factor analysis as described above. A similar approach was taken in the Motivation At Work Scale [29]. As Gagne et al [29] point out, distinction between these two factor has been difficult based on psychometric properties of SDT-based scales. Thus, a 21-item scale based on four subscales was finally developed.

The scale is designed to be self-administered. There is no summary score for the entire scale, because each hypothesized factor is viewed as a distinct type of motivation. Moreover, according to the SDT, when predicting different outcomes it is not the overall amount of motivation but the type of motivation that matters [19]. Each of the four

subscales, based on the respective factor, can have its own score which is an arithmetic average of the items in that subscale.

Validation variables

Validation variables were constructed only for the RBF pilot sample (n=805) as they were used only at the final validation stage of the scale (see the last level of boxes in Figure 3.2).

Intention to quit was constructed based on the originally continuous variable. Given the binary nature of the decision (stay or quit) it appeared that a binary variable would be more appropriate than a continuous one. Health workers were asked how long they intended to stay in this facility and had to mark years, months or days. The original continuous variable ranged from 0 days to 60 years. However, it had clear outliers, first identified using box plots and more formally with "lv" function in STATA [103]. Values above or below the inner fences are considered to be mild outliers and values above or below the outer fences are considered to be severe outliers. Thirteen severe outliers identified in this way were assigned the value of the inner fence, which was 22 years. Of the 796 health workers that answered this question, it appeared that the majority (65%) intended to stay in one facility for 5 years or less, of which 35% intended to stay 1 year or less. Based on this, we redefined the variable as a binary variable where those who intended to stay 1 year or less (n=181) were defined as "intend to quit" group.

The performance assessment and feedback variable was constructed based on two variables. Health workers were first asked whether or not they had received a formal employee performance assessment in the past 12 months, to which they could answer "Yes", "No", "Don't know". Of the 804 health workers who responded to this question, 6% said "Don't know". These were combined with those who said "No" (12%) because (a) it seems that the intended benefits of having a performance assessment disappear if the employee is not aware of it, and (b) both groups, those who said "No" and those who said "Don't know", were instructed to skip the following question regarding the feedback. Those who responded "Yes" to the question on assessment were asked whether or not they had feedback from that assessment. Of the 648 health workers who answered this question, 76% said "Yes", while 24% said "No". These two variables were combined into one binary variable as it was regular feedback that was hypothesized to have a positive impact on motivation. Thus, the new variable was coded as "1" if a health worker received assessment and feedback, and "0" if she received assessment but not feedback, or did not receive assessment at all. Of the 804 health workers, 494 (61%) received assessment and feedback.

Staff voice or participation was measured through a question: "Staff in this facility have opportunities to express their opinions". Fairness of the supervisor was measured through a question: "My supervisor is unfair to me". Equity of rewards was measured through a question: "I feel like I am rewarded fairly for the work I do". Responses to all three questions were measured on a four-point Likert scale (1="Strongly disagree", 2="Disagree", 3="Agree", 4="Strongly agree"), the same response scale that was used

for the motivation scale. Of the 804 health workers that responded to the question on opportunities for staff to express their opinion only 6% said "Strongly disagree", while 46% said "Strongly agree". Of the 803 health workers that responded to the question on fairness of the supervisor, 12% marked "Strongly agree", while 46% expressed the opposite opinion. Of the 803 health workers that responded to question on equity of rewards, 12% marked "Strongly disagree", while 28% marked "Strongly agree". All three questions were part of the satisfaction scale administered to health workers together with the motivation scale. However, as the psychometric properties of the satisfaction scale were not assessed yet, these three items were used separately for the validation of the current scale.

General motivation was constructed based on one item "Overall, I feel very motivated to do my job", measured on a four-point Likert scale similar to items on the work motivation scale. It was treated as an observed categorical variable. Of the 797 health workers that answered this question, less than 1% expressed a strong disagreement with the statement. Almost 98% of the respondents marked either "Agree" or "Strongly agree".

3.4 Data analysis

3.4.1 Missing data

None of the items in the work motivation scale had more than 0.5% missing values. The variable on general motivation had 5.3% missing values. The missing values for the question on general motivation were almost exclusively restricted to areas where

community survey method was used due to security considerations (92.2%) and in provinces outside of the RBF pilot (93.8%). None of the other variables, except intention to quit, had more than 5% missing values. Intention to quit had 6.2% missing values. Similar to general motivation variable, 78.7% of these were from facilities that were surveyed through community survey method (see Chapter 2). Moreover, they were almost exclusively restricted to provinces outside of the RBF pilot (94%). These variables were used for validation only with the RBF pilot sample; thus, the proportion of missing values relevant for the purposes of the study was less than 2%.

Given that scale items were measured on ordinal scale, a robust weighted least squares estimator was used for model estimation. With the WLSMV estimator without covariates in the model, the pairwise present method is used in MPlus [104]. As the name implies, in pairwise present analysis missing data is looked at for pairs of variables. For example, if 100 respondents have non-missing values for y1 and y2, 100 observations are used to compute the correlation between y1 and y2. If 90 respondents have non-missing values for y1 and y3, 90 observations will be used to compute the correlation between y1 and y3. As proportion of missingness was less than 2% on observed covariates that were used as validation variables in this study and missingness is not allowed for the observed covariates [104], observations with missing data on covariates were deleted.

3.4.2 General approach

The data analysis proceeded in several distinct stages. First, direction of response for each item, particularly for those that were negatively worded, was checked to ensure

correct understanding of items by the respondents. Next, mean and median scores, proportion of missing values, and frequency distributions of responses to individual questions were examined. In RBF pilot facilities only 4.8% used Pashto forms, while in facilities outside of the RBF pilot 45.9% used Pashto forms. Given that reliability and validity of the scale can differ by language, the analysis of non-RBF data sets was restricted to Dari forms only. Thus, Pashto forms in non-RBF pilot facilities were excluded from the EFA and CFA conducted on non-RBF data sets. The Pashto data set was kept separately and used at the final stage for assessing the scale for generalizability in order to ensure that it would be useful for measuring motivation not only among Dari speakers, but also among the Pashto speakers (Figure 3.2).

According to Raykov and Marcoulides [84], CFA is a method of testing well-developed hypotheses about the latent variable structure as the number of factors is stated in advance, and for each indicator it is pre-determined which of its loadings will be 0 on which factor. Although in this study we had an initial hypothesis according to which motivation was a multi-dimensional construct, it was determined that exploratory analysis would be more appropriate as the first step, as the theory underlying the construct and existing instruments was derived from a very different social context.

According to Raykov and Marcoulides [84], a "highly beneficial application of CFA is as a follow-up to EFA" on the same set of observed variables but using a different data set derived from the same studied population. Thus, we split our sample into three parts. The Dari-only data set containing facilities outside of the RBF pilot was randomly split into

two parts (further referred to as Group 1 non-RBF and Group 2 non-RBF). Exploratory factor analysis was performed on Group 1. Based on these results, confirmatory factor analysis was performed on Group 2 data [101]. Finally, RBF facility data was analyzed using the CFA model that resulted from Group 2 non-RBF sample.

Reliability and validity measures were assessed on the final subscales. Standard error computations used a sandwich estimator to account for non-independence of observations as a result of cluster sampling. No sampling weights were applied. All data were analyzed using STATA 12 [105] and MPlus v7 [104].

3.4.3 Factor analyses

An exploratory factor analysis using the oblique GEOMIN rotation was conducted on the 21 items of the work motivation scale on Group 1 non-RBF sample. The factors are correlated under the oblique GEOMIN rotation, which is in line with our theoretical understanding of motivation factors. Given that responses were measured on ordinal scale, the model was fit to a polychoric correlation matrix using the method of robust weighted least squares [104]. Standard error computations used a sandwich estimator in order to account for non-independence of observations due to cluster sampling by specifying TYPE=COMPLEX in the ANALYSIS command of the input file [104].

As the preliminary step to EFA the model was checked for specification errors, identification problems and input errors by examining the results for negative error variances, correlations greater than one, and extremely large parameter estimates[106].

To determine the appropriate number of factors to retain several criteria were used. First, Kaiser-Guttman criterion was used (eigenvalues>1). This was coupled with evaluation of the proportion of total variance that was explained by given factors. Parallel analysis was not applicable as the data was measured on an ordinal scale [104].

Next, model fit statistics, factor loadings, communalities, and factor structure were examined. Given the sensitivity of the chi-square test to sample size, the comparative fit index (CFI), Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean residual (SRMR) were used to evaluate the global fit of the model. Use of several indices to judge the model fit is recommended as each of them has its strengths and weaknesses [107-109]. CFI ranges from 0 to 1, with 1 indicating a very good fit. Typically, CFI>0.9 indicates acceptable fit [110], although Hu and Bentler suggest 0.95 as a cutoff [108]. TLI is usually lower than CFI, but it has similar cutoff level where an index<0.9 indicates that a model needs to be re-specified [110]. RMSEA <0.08 indicates a reasonable fit, although <0.05 is desirable [111]. Similar to RMSEA, smaller value for SRMR (0.08) indicates that a model may be considered to provide a good fit for the data [108].

Items with factor loadings higher than 0.30 were retained if they loaded clearly on one of the factors [84]. As Raykov and Marcoulides explain [84], when loading is less than 0.30, one particular factor explains less than 10% of the observed variance in the item. Thus, measures with loadings higher than 0.30 are “considered as generally contributing to factor interpretation” [84]. In addition, modification indices (MI) were examined.

Modification indices are useful in evaluating assumptions such as non-correlated residual errors by giving the researcher a sense of what happens when those assumptions are relaxed [112]. Ideally, the MI values should be less than 3.84 (the critical value under a chi-square distribution with 1 df), but 10 is still considered to be a conservative value and is a default option in MPlus [104]. Anything above 10 is the evidence that the underlying factor does not explain all the covariance between items that have high residual correlation, so there may be a need for additional factors or these particular items do not work.

Based on the results of the EFA, CFA was fit on Group 2 non-RBF sample. As described in the case of EFA, given that responses were measured on ordinal scale, the model was fit to a polychoric correlation matrix using the method of robust weighted least squares [104]. Standard error computations used a sandwich estimator in order to account for non-independence of observations due to cluster sampling by specifying TYPE=COMPLEX in the ANALYSIS command of the input file [104]. All factor loadings were freed, and thus, the metric of the factors was defined by fixing the factor variances at one.

All of the above described fit indices, except SRMR (not provided for CFA models with complex data), were used to evaluate the global fit of the CFA model. Values of the standardized factor loadings as well as their statistical significance were examined. In addition, factor correlations were examined for consistency with theoretical predictions of the self-determination theory on motivation. Finally, r-square estimates that show the

variance explained in the latent response variable underlying the categorical variable were examined to determine quality of the items in the model. The final CFA model was then tested on RBF pilot data set.

3.4.4 Reliability

According to DeVellis [113], scale reliability is "the proportion of variance attributable to the true score of the latent variable". Item-total correlation omitting the item in question from the total as described by Raykov and Marcoulides [84] is a basic useful measure of reliability that was applied in this study.

Internal consistency based on Cronbach's alpha coefficient is one of the key measures of reliability. In simple terms, items should be correlated with each other and the total score. While Cronbach's alpha coefficient is a convenient and relatively simple measure of reliability, it is vulnerable to outliers, non-normality of data, small number of items in the scale, and low variability in total scores [114, 115]. As the current scale used a four-point Likert scale, reliability for each subscale was estimated using ordinal alpha based on polychoric correlation matrix as per Gadermann, Guhn & Zumbo [100]. Internal consistency coefficients such as Cronbach's alpha can be modified and estimated using polychoric correlation matrix [100, 116].

3.4.5 Validity

Validity concerns the extent to which the proposed scale measures the true construct. According to DeVeillis [113], "validity is inferred from the manner in which a scale was

constructed, its ability to predict specific events, or its relationship to measures of other constructs". Several types of validity measures were used in this study. The summary of the variables used for construct validity, hypotheses and rationale are provided in Table 3.14.

First of all, an indication of content validity is the process of scale construction, including thorough review of the theory of motivation and existing scales, described earlier. The focus group discussions conducted as part of field testing of the instrument also support the content validity of the scale. Second, cross-validation of the model is considered as one of the important measures of validity [106]. Following Bagozzi and Yi [106], as part of the cross-validation process, parameters of our model were estimated in one sample (Group 2 non-RBF), and the predictive effectiveness of the model was subsequently determined on a separate, independent sample drawn from the same population as the first (RBF sample). The model with the best fit to one data set may capitalize on chance or specific characteristics of that data set, but a valid model should fit a new sample as well [106].

Construct validity was measured through the differentiation between groups [84]. The variable used for this analysis was intention to quit. It was hypothesized that there would be differences in factor means between those intending to stay and those intending to quit [117, 118]. Specifically, it was expected that the factor on amotivation would have a lower mean in those intending to stay, thus signifying that amotivation was higher among

those intending to quit. The analysis was conducted using the mixture modeling with known classes approach as described in MPlus User's Guide [104].

Performance assessment and feedback was used as predictor of motivation factors. According to SDT, individuals have three innate psychological needs, one of which is the need for competence [119], and meaningful performance feedback promotes a sense of competence and autonomy when people feel responsible for their successful performance [19, 120]. A positive association was expected between this variable and identified regulation factor, indicating that health workers who received regular performance assessment and feedback would have higher identified regulation, as predicted by the theory. The analysis was conducted using structural equation modeling where outcome variables were continuous factors measured through observed categorical indicators [104].

Variables on staff voice and equity were used to test construct validity through the latent variable modeling framework where staff voice and equity variables were predictors of motivation factors. Staff voice was expected to have a positive association with identified regulation as it would help the staff feel that their managers care about their point of view and acknowledge their perspectives, which would increase the sense of competence and autonomy among employees [120, 121]. SDT is not as clear about the role of staff voice in introjected motivation. However, to the extent that staff voice as ability to express opinions increases one's self-esteem and promotes the feeling of self-worth, it is reasonable to expect a positive association. Lack of fairness of the supervisor

was predicted to have a negative association with identified regulation and a positive association with amotivation as predicted by the theory of motivation focusing on organizational justice as well as SDT [19, 54, 122]. Equity of rewards, one of the main elements of organizational justice, was expected to have a positive association with identified regulation, external regulation and introjected motivation factors [122].

Construct validity was further evaluated using the question on general motivation. Hansen [39] in his dissertation proposed that validity of a scale on patient satisfaction can be measured by assessing the association between the scale and a related variable, such as the overall patient satisfaction. Using this approach, this study measured construct validity by examining the strength of association between the proposed scale and the item on overall motivation. The formula provided by Raykov [123] for continuous factor indicators was modified to fit the case of data measured on ordinal scale. Thus, we used ordinal alpha estimates for reliability instead of the reliability estimation process described in Raykov to correct the correlation between the criterion and the factors [123]. Polychoric correlation coefficients of the summed scores for each factor and general motivation were also applied as an additional check.

3.4.6 Measurement invariance for assessing generalizability to Pashto language

Measurement invariance is an important characteristic, particularly when the scale is intended to be used in different linguistic or cultural groups. If the measurement equivalence is not established across two languages or countries, it is hard to make any

comparisons between the two groups, as it is unclear whether the interpretation of the scale items is the same [124, 125]. The test of the measurement invariance of the work motivation scale between Pashto and Dari languages was conducted using the following three step procedure [104, 106, 124]:

1. Configural invariance was assessed using CFA based on the factorial structure of the final RBF model;
2. Metric equivalence was assessed by imposing equality constraints on factor loadings;
3. Invariance of thresholds was assessed while holding factor loadings and thresholds constant across groups.

Satisfactory model fit indices for the three models with sequentially imposed constraints, beginning with the least restrict one, would provide evidence to the generalizability of the scale developed based on Dari language speakers to Pashto speakers in Afghanistan.

3.5 Results

3.5.1 Study population

A total of 2,413 health workers drawn from 775 facilities (clusters), covering 34 provinces of Afghanistan, were surveyed in this study. Of these, 94% were surveyed using professional survey teams, while 6% were surveyed using community surveyors (Table 3.2). Community survey method was used in the most insecure areas and largely restricted to areas outside of the RBF pilot provinces. More than 50% of the sample came from BHCs. The majority of surveyed health workers came from facilities either managed by MOPH with support of national or international institutions (53%) or by

NGOs only (42%), and only a small portion (5%) came from facilities managed by MOPH only.

Approximately two-thirds of the sample used Dari as their main language. Given the reality of gender distribution of health workers in Afghanistan, 64% of surveyed health workers were male. Vaccinators made the largest group of health workers in our sample (29%), followed by nurses (24%), and then doctors (14%) (Table 3.2). Almost 60% of the sample could be considered to be fairly new employees as they had worked in their current facility for three years or less. Only 6% of surveyed health workers reported that they had worked in the same facility for more than 10 years. When asked about their intention to stay in the current facility, 28% stated that they intended to leave the facility within a year.

The majority of health workers stated that they received supervisory visits 30 days preceding the survey (80%) and formal employee assessment in the past 12 months (83%). Of those who received employee assessment, 82% received feedback from it. Only two-thirds of the respondents received their salaries on time over the past year, and slightly more than one-third received a salary increase over the same period (Table 3.2). Eighteen percent stated that they received performance-based payments. Eighty percent were estimated to have a second job. One third of the sample or 805 health workers drawn from 256 facilities, covering 11 provinces, were part of the RBF pilot program.

3.5.2 Descriptive results for work motivation scale

Means, standard deviations, and sample sizes for each individual item of the original work motivation scale are presented separately for each of the three validation groups

(Appendix A, Table 7.3). It also presents marginal frequency distributions for each item by validation group. Items that were selected for the final scale are bolded. Four items on amotivation are negatively worded items and they were not re-coded because leaving them in the original scale made the interpretation of loadings and regression coefficients more intuitive.

The means for items for all three groups were generally high, above 3.0 on a 4-point scale, except the four items on amotivation described in further detail below. For non-RBF Group 1 and Group 2 samples, only three items (items No. 9, 12, 16) had a mean below 3.0: family, pay, and benefits. In addition to these, the RBF sample had one more item with a mean below 3.0 (item No. 10), respect in the community. Looking at frequencies of endorsement, out of 21 items of the scale, only three items had more than 10% of respondents who marked "Strongly disagree" to positively worded questions in non-RBF Group 1 and Group 2. These are the same items that had means below 3.0. In the RBF sample, in addition to these three items, item No. 10 (respect in the community) and No. 14 (sufficient resources) showed more than 10% of respondents expressing strong disagreement with the statement. These items are generally considered to be strong items.

On the opposite end, for all three samples item on God's grace as the motivating factor had the highest frequency of endorsement with less than 5% of respondents marking either "Strongly disagree" or "Disagree". Moreover, this was the only item that had an endorsement frequency of approximately 90% in "Strongly agree" category. Streiner and

Norman [126] in their practical guide to the development and use of scales advise against items with frequency of endorsement above 95% or below 5%. Even keeping items with frequency of endorsement between 5% and 95% will lead to highly skewed items. At the same time, it is recommended not to throw out items simply based on this criterion [126]. Thus, the item on God's grace that was introduced into the scale based on the results of the focus group discussions and further analysis of Afghan culture where religion plays an important role in daily life was not eliminated from the scale.

In general, satisfaction and motivation data tend to be skewed in a positive direction. For example, in the Patient perception of quality and satisfaction scale developed and used in Afghanistan [39], none of the items had a mean below 3.0. Moreover, the frequency of endorsement for category 1("Strongly disagree") was less than 5% for all items, except one item on the cost of the visit. The scale was measured on a four-point Likert scale and thus comparable to the response scale in this study.

3.5.3 Exploratory factor analysis

Results from the initial exploratory factor analysis conducted on 21 items indicated that one factor explained only 25% of the total variance. Four factors explained 52% of the total variance, which increased slightly to 58% with five factors (Table 3.3). Kaiser-Guttmann Criterion suggested retention of five factors, as the eigenvalue of the fifth factor was 1.1. However, this criterion has been criticized, as it often overestimates the number of factors to be retained [127]. As the next step, the global fit indices were examined. Based on these, one- and two-factor models were rejected (Table 3.4). A three-factor model had satisfactory fit indices but it would explain less than 50% of the

total variance. Together, these three criteria suggested that models with four or five factors would provide the best fit to the data. Factor loadings, unique variance estimates and modification indices of the four- and five-factor models were examined next.

In a four-factor, model items “I work in this job because I have a chance to help other people through my work” (intr1), “I feel I should personally take the credit or blame for the results of my work on this job” (intr3), “I feel a very high degree of personal responsibility for the work I do on this job (intr7)”, “I work in this job to gain God’s grace” (intrj4) loaded clearly together on one factor with positive loadings of 0.50 or higher and significant at 5% level. This factor could be termed an identified regulation factor (Table 3.5).

Items “I work in this job because it allows me to use my skills” (intr5), “I work here because of opportunities for promotion” (extrg2), “I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)” (extrg3), “I work here because it is located in a safe area” (extrg4), “I work here because of good benefits I receive (Note: all benefits – housing, transportation, anything else you receive – think overall)” (extrg5), “I work here because it provides long term security for me” (extrg6) – all loaded clearly on one factor with positive loadings ranging from 0.43 to 0.71. While loadings varied in size, they were all above 0.30 and were statistically significant. This factor seemed to form external regulation factor.

Items “I work here because it makes me feel important” (intrj1), “I do this job because my family would be disappointed if I quit” (intrj2), “I do this job because it gives me respect in the community” (intrj3) – all loaded together with positive loadings (p-value <0.05) ranging from 0.40 to 0.80. Item “I only work here to get paid” (extrg1) had a positive statistically significant loading on this factor but it was marginally acceptable as the common criterion for acceptable loadings as described in the Methods section was higher than 0.30. Looking at factor structure (Table 3.6), this item did not seem to have an observably high correlation (positive or negative) with any of the four factors. The highest correlation was with factor 3 (0.37), but it was still below correlations of other items in this factor. Also, it had the highest residual variance of 0.77 in the entire scale. Finally, it was conceived as an external regulation factor item, as it focused on pay but seemed to load with items conceived as introjected motivation factors (self-esteem, guilt). For these reasons, this item was not included in this factor.

Items “I frequently think of quitting this job” (amot1), “I sometimes feel my work here is meaningless” (amot2), and “I don’t care much about the quality of work here” (amot3) – all had negative statistically significant loading coefficients on identification factor. These loadings ranged from -0.33 to -0.62. However, items amot1 and amot2 also loaded together on a separate factor with positive loadings of 0.46 (p-value<0.05). Residual variance for amot3 was below 0.70 and approximately equal to residual variances of amot1 and amot2. Moreover, amot3, as the other two items in this factor, was originally constructed as an amotivation item. It also had a high, although negative, loading on identification factor as did amot1 and amot2 (Table 3.5). Finally, it had very low

modification indices (<3.84). Based on these multiple pieces of evidence, it seemed that this would be an item worth keeping, particularly as the factor on amotivation would otherwise have only two items.

Other items in the scale "I work in this facility because it plays an important role in the community" (intr2), "I work in this job because it allows me to decide how my work is organized" (intr4), "I work in this job because I can accomplish something worthwhile in this job" (intr6), "I work here because I have no other choice" (amot4) had loadings of similar size on more than one factor, although they were above 0.30. Also, item amot4 had one of the highest residual variances and the three highest modification indices (14.0, 10.2 and 7.9).

Many of the items that loaded together in a four-factor model continued to have significant loadings on the same factors with a five-factor model. Items intr1, intr3, and intrj4 loaded together, but item intr7, which loaded on identified regulation factor previously, had similar loadings on identified and external regulation factors. Items intr5, extrg2, extrg3, extrg4, and extrg5 loaded together on one factor. Two of the items on introjected motivation (intrj1 and intrj3) also loaded together, while intrj2 showed similar loadings on this factor and a new factor, consisting of extrg1, extrg6 (previously loading on external regulation factor), and amot4. This new factor however, was largely driven by amot4 that had a loading of 0.75, almost twice the size of other loadings in this factor. The factor on introjected motivation (intrj1 and intrj3) had an additional item intr2 ("I work in this facility, because it plays an important role in the community, but one of the

three items (intrj1) had an extremely high loading of 0.96. These types of extreme differences in the size of factor loadings in one factor should generally be avoided. For example, items with large differences in loadings will violate one of the assumptions for estimating Cronbach's alpha [84].

Also, with the exception of two items (intrj1 and amot4), communalities did not observably increase when the fifth factor was added to the model. Thus, it seemed that residual term variances provided evidence that proportion of unexplained variance for most of the items in the scale did not improve with the additional factor. Given these results, it seemed reasonable to accept a four-factor model containing the following four factors: identified regulation, external regulation, introjected regulation and amotivation. This factorial structure is in line with the initial hypothesis based on SDT, although the two autonomous types of motivation, integrated and identified regulation, came out as a single factor. The structure of the scale reflects the view of extrinsic motivation as a spectrum of different types of motivation, reflecting the degree of self-determination or autonomy.

3.5.4 Confirmatory factor analysis

Confirmatory factor analysis on non-RBF Group 2 sample was conducted to confirm the hypothesized factorial structure of the scale based on the results of the EFA conducted on non-RBF Group 1. A 16-item four-factor model (Model A) tested using CFA on non-RBF Group 2 sample resulted in less than satisfactory fit indices (Table 3.7): CFA of 0.87, TLI of 0.84, RMSEA of 0.06 (90% CI: 0.05; 0.07). Based on the review of factor

loadings, it appeared that item extrg5 on good benefits should be dropped from the model as it had the loading of 0.21, well below the standard. Having eliminated this item from the scale, the analysis was repeated on a 15-item four-factor model (Model B). This resulted in good fit indices with all standardized factor loadings above 0.30 (Table 3.7 - Table 3.8).

Overall, factor correlations were also in the expected direction based on the theoretical framework underpinning the scale (Table 3.9). Identified regulation factor had significant positive correlations with external regulation and introjected regulation factors (0.61 and 0.49, respectively), and a significant negative correlation with amotivation factor (-0.69). External regulation factor had a strong positive correlation with introjected motivation factor (0.68) and a negative correlation with amotivation (-0.50). However, one unexpected result was a lack of correlation between introjected motivation and amotivation factors.

The four-factor model with 15 items described above (Model B) was finally tested on the RBF pilot sample. The global fit indices for the model were satisfactory: CFI of 0.93, TLI of 0.91, RMSEA of 0.05 (90% CI: 0.04, 0.06) (Table 3.10). All loadings were well above the cut-off of 0.30 and r-square estimates that show the variance explained in the latent response variable underlying the categorical variable were above 0.30 for most of the items (Table 3.11).

Similar to results of the CFA model on non-RBF Group 2, identified regulation factor had significant positive correlations with external and introjected regulation factors (0.68 and 0.31, respectively), and a significant negative correlation with amotivation factor (-0.60) (Table 3.12). It should be noted here, however, that the size of the correlation between identified and introjected regulation was in this case lower (0.31 vs. 0.50). External regulation factor had a strong positive correlation with introjected regulation factor (0.72) and a negative correlation with amotivation (-0.27). The negative correlation with amotivation factor was lower in RBF as compared to non-RBF sample. Also, introjected motivation factor had a positive, although small, correlation with amotivation (0.28).

3.5.5 Reliability

Inter-item and item-total correlations were examined for reliability for each of the sub-scales (Table 3.13). For all four sub-scales, item-total correlations estimated omitting the item in question from the total score were generally high, well above the cut-off level of 0.20 suggested by Streiner and Norman [126]. For identified regulation factor, they ranged from 0.33 to 0.45, where item intr3 ("I feel I should personally take the credit or blame for the results of my work on this job") had the lowest correlation. For external regulation factor, the item-total correlations ranged from 0.35 to 0.46, where item extrg4 ("I work here because it is located in a safe area") had the lowest correlation. For introjected regulation factor, the correlation coefficients ranged from 0.32 to 0.47, where item intrj2 ("I do this job because my family would be disappointed if I quit") performed

least well. The factor on amotivation had the highest item-total correlation estimates (0.45-0.55).

Internal consistency reliability was measured separately for each of the four subscales formed on the basis of a final four-factor model, as described in the preceding section (Table 3.13). As responses were measured on a four-point Likert scale and were highly skewed, an ordinal alpha based on polychoric correlation matrix was used instead of traditional Cronbach's alpha. Ordinal alpha coefficients for the work motivation subscales ranged from 0.64 for introjected regulation factor to 0.75 for identified regulation factor. These values can be considered satisfactory as discussed in section 3.7. Ordinal alpha coefficient was not estimated for the entire scale, because like Cronbach's alpha, it assumes unidimensionality of the set of components of a given scale and this scale was specifically designed to not have this characteristic [115].

3.5.6 Validity

Results of the cross-validation of the model as a measure of validity are discussed above as part of the CFA results. As described, the model was developed based on one sample (non-RBF Group 1). Next, the model was fit on non-RBF Group 2 sample ("calibration sample") [106]. The model derived from the calibration sample was then used on a new sample (RBF pilot group). According to Bagozzi and Yi [106] fixing all parameters in the second sample, including factor loadings, to parameters derived in the calibration sample is a very strict requirement. Fixing factor pattern was determined to be a sufficient evidence of the validity of the scale. This is a less stringent level of cross-validation, and

only the factor pattern is validated across calibration and validation samples [106, 125]. The model applied to the RBF pilot sample had good fit indices, providing evidence of configural equivalency. Furthermore, the magnitude of factor loadings between the two samples changed very little.

Construct validity was measured through the differentiation in motivation factor means between those intending to quit with the next 12 months and those intending to stay (Table 3.15). Multiple group analysis was performed using MPlus mixture modeling with known classes approach [104]. As expected, the factor mean for amotivation was significantly lower (p-value<0.05) in the group that was intending to stay, as compared to the group that was intending to quit. In other words, amotivation was higher among those intending to quit within a year or less.

The performance assessment and feedback variable was used as a predictor of motivation factors. None of the coefficients, except the coefficient on introjected motivation factor, was statistically significant. Moreover, they were close to zero. The variable was a simple binary variable that did not reflect frequency or quality of either performance assessment or feedback following it, which may explain these results [120].

Ability of the staff to express their opinions (staff voice) had statistically significant (p-value<0.05) and positive association with identified and introjected regulation factors (Table 3.16). This was in accordance with SDT and the hypotheses as described in Table 3.14. Association of staff voice with external regulation was unexpected. SDT is no very

clear about the expected relationship between external regulation and variables such as staff voice and participation. But if staff voice is about promoting the feelings of autonomy and choice, it is expected to be associated more with autonomous and less with controlled types of motivation [120]. The significant negative association with amotivation was too small (-0.09) and therefore cannot be taken as evidence of the relationship between staff voice and amotivation.

As expected, unfairness of the supervisor had a strong negative association with identification (p-value<0.05), and a strong positive association with amotivation (p-value<0.05) [122]. Equity of rewards appeared to be a strong predictor of external regulation and introjection (p-value<0.05), also confirming the hypothesis and providing evidence of the validity of the work motivation scale. Equity of rewards had no association with identification. While it was reasonable to expect a positive association between the two variables if one viewed equity of rewards as one of the components of organizational justice [122], this is just one of its elements and perhaps, not the most important one, for health workers with strong identified regulation. These health workers, even if they feel they are not being fully rewarded for the work that they do, might still feel strong motivation for their work.

Construct validity was further evaluated using the question on general motivation. Polychoric correlation between motivation factors and the item on overall motivation corrected for reliability of the subscales provided further evidence of the validity of the proposed scale (Table 3.17). Identified regulation factor had the highest correlation

(0.59), followed by external regulation factor (0.36). Amotivation had a negative correlation (-0.30), providing evidence of discriminant validity. The lowest correlation coefficient was with the introjection factor (0.11). Polychoric correlation coefficients of the summed scores for each factor and general motivation were also applied as an additional check and gave very similar in value (not shown here).

3.5.7 Measurement invariance for assessing generalizability to Pashto language

Measurement invariance was tested by sequentially imposing stricter constraints beginning with the least restrictive model. As the first step, the four-factor 15-item model based on non-RBF Group 2 and RBF pilot samples was fit on the non-RBF Pashto Group. The Model 1 (Table 3.18) had satisfactory fit indices, indicating that the first threshold for invariance was met and stricter constraints could be imposed. Next, the loadings between the two groups were constrained to be equal. The fit of the model remained satisfactory based on global fit indices. The invariance of thresholds while holding factor loadings constant across groups was tested as the final step. The fit indices were less than satisfactory for this model with CFI falling below 0.90. Thus, metric invariance based on equal factor loadings was the highest level of invariance that would still provide reasonable fit to the data. If the loadings are same and the factor means are fixed at 0 for both groups, then the true score has to be the same for both groups. In this case, difference in thresholds means that for the same underlying true score, respondents in different language groups would respond differently. This provides evidence that the

number of latent dimensions of motivation (motivation factors) and its structure (e.g., the factor pattern matrix) are equivalent across the two language groups.

3.6 Discussion

Health outcomes depend on the availability and performance of the health workforce, which is one of the six building blocks of a health system [128]. Motivated health workers are less likely to leave their jobs and contribute to a shortage of health workers, particularly in poor rural areas. Moreover, studies indicate that motivated health workers, specifically those with more intrinsic type of motivation, provide higher quality care, even in resource-poor settings and in circumstances where they are severely underpaid or not paid at all [77, 92]. Thus, health workforce motivation is one of the key objectives of health system interventions such as performance-based financing. However, it has proven difficult to provide evidence of the effectiveness of such interventions on improving health worker motivation and even harder to link that change to the change in performance, because of the measurement challenges related to motivation. Therefore, the multidimensional work motivation scale developed and validated in this study is a highly relevant tool for policy-makers and practitioners, particularly in developing countries such as Afghanistan.

The scale is based on a clear theoretical framework that provides a more subtle view of motivation and does not draw a sharp distinction between intrinsic and extrinsic motivating factors. Instead, it takes a more nuanced approach to understanding different dimensions of motivation based on the Theory of Self-Determination. According to SDT,

different dimensions of motivation are associated with different organizational factors and have different consequences on performance [19]. For example, to promote more autonomous types of motivation, such as identified regulation, managers should implement interventions that would allow vertical and horizontal enlargement of the job, institute mechanisms for providing regular autonomy-supportive feedback for employees, and involve staff in decision-making processes. Motivation among employees who have mainly introjected regulation maybe strengthened by increasing their status within the organization and community perhaps by changing their job titles or providing community recognition for their work. These interventions do not require substantial financial resources unlike interventions that focus on extrinsic rewards such as bonus payments. This is particularly important in resource-poor settings such as Afghanistan.

The scale is short with only 15 items, but at the same time, it reflects the multidimensional view of motivation described above. The initial hypothesis regarding the factorial structure was confirmed, although the items on integrated and identified regulation had to be combined based on the results of the initial exploratory factor analysis supported further through confirmatory factor analysis. This was not an unusual approach as described earlier in the paper as distinction between these two factors was problematic in other studies [29].

The scale demonstrated good psychometric properties. First of all, it performed well on two reliability measures. It showed item-total correlations well above the rule of thumb of 0.20 suggested by Streiner and Norman [126]. Three out of four subscales had ordinal

alpha of 0.70 or higher. The introjection subscale with the lowest internal consistency coefficient among the four (0.64) still demonstrated higher reliability than subscales on comparable measures. Cronbach's alpha for similar types of scales in other studies had a range from 0.36 to 0.92. The coefficient alpha tended to be even lower when applied to subscales. The General Causality Orientation Scale, using a seven-point Likert response scale, reported 0.77 and 0.59 for autonomy and control orientation subscales, respectively [129]. The Work Extrinsic and Intrinsic Motivation Scale (WEIMS), that also used a seven-point Likert scale, had alpha values ranging from 0.64 (amotivation) to 0.83 (integration) [23]. The authors of the study argued that these values suggested adequate reliability, particularly since there were only three items per subscale. The original French WEMS alpha values ranged from 0.59 to 0.77 [23]. Subscales in the health worker motivation scale used in Kenya [86] had Cronbach's alphas ranging from 0.36 to 0.64. The Scale had 3-5 questions per subscale, similar to the scale presented here. Systematic Review of Job Satisfaction Scales [130] reported internal consistency measures ranging from 0.52 - 0.92 for total scales and subscales.

Several measures of validity were applied in this study. First of all, using CFA, the model developed in one sample was tested on a separate sample. While not all of the parameters remained stable across the two samples, the stability of the overall factorial structure provided evidence of the validity of the scale and the predictive effectiveness of the model. Secondly, discriminant validity was tested and supported by the significant negative correlation found between general motivation question and amotivation. Large significant positive correlation with identification provided further support for the

construct validity of the scale. Correlations that were smaller in magnitude, although still significant, with external regulation and introjection, indicate the need for a scale with multiple items that measure different aspects of motivation. These findings reflect the multidimensional nature of motivation that is not adequately captured by a single item. Thus, the general question on motivation may not capture the different dimensions of a construct such as motivation quite as well as a scale consisting of multiple items and sub-scales.

Secondly, the results of the multiple group analysis demonstrating the difference in factor means for amotivation between respondents intending to quit and those intending to stay provided evidence of the construct validity of the scale. Health workers who were intending to quit within the coming year were significantly more amotivated than their colleagues. Association between turn-over or intention to quit and poor motivation has been demonstrated in other studies [29, 131, 132].

Although it was expected that performance assessment and feedback would be positively associated with identified regulation as predicted by SDT [120, 133], this study did not find any association between these two variables. This may be due to the fact that the variable representing assessment and feedback was a simple binary variable which did not indicate anything about the quality of these processes. In a qualitative study on health provider motivation in Tanzania, the respondents appeared to attach little value to the feedback from their supervisors, as it was rarely helpful or supportive [90]. How

assessment and feedback are provided plays a vital role in whether they have any impact on worker motivation [54, 120].

As expected, staff voice had a positive association with identified and introjected regulation factors. The ability of staff to express their opinions and participate in decision-making is considered to be one of the determinants of staff motivation and commitment to organizational goals [19, 133]. As hypothesized, unfairness of the supervisor had a strong positive association with amotivation and a strong negative association with identification. According to the organizational justice theory of motivation and its further elaboration in SDT, perception of fairness is a key element in fostering autonomous types of motivation, including identification [19, 54].

Equity of rewards appeared to be a significant predictor of controlled types of motivation (external and introjected regulation). While the expectation was that this item as it relates to equity would also be associated with identification, there seems to be a reasonable explanation as to why this part of the hypothesis was not confirmed. This variable was very specific and asked only about equity of external rewards. It is likely that health workers whose main motivation was more intrinsic were less concerned with equity of rewards than their colleagues who were primarily extrinsically motivated. Overall fairness of the supervisor seemed to be of greater importance to those who were more intrinsically motivated.

Finally, the results of the measurement invariance test provided sufficient evidence that the scale could be applied to Pashto speakers as well as Dari speakers in Afghanistan. This is an important finding given that Pashto is the second official language after Dari and Pashtuns make up the largest ethnic group in the country [134].

There are three limitations to the present study. First of all, a more in-depth qualitative study would provide a richer understanding of the construct of motivation in Afghanistan and could improve the quality of the scale. The underlying theoretical framework is based on Western theories of motivation, behavior and organization. Cultures where maintenance of status quo, propriety, respect for tradition, familial obligations and respect for authority are strongly emphasized, the key assumptions of the motivation theories that provide the basis for the models developed for this study may not be appropriate [26]. However, this limitation underlies most work that is done outside of the USA and Europe on work motivation, although there are a few studies that have tried to explicitly incorporate such cultural differences [47]. Moreover, the results of the focus group discussions and the validity measures of the work motivation scale developed in this study suggest that motivation defined as a process that explains the "direction, amplitude, and persistence of an individual's behavior" applies across cultures [13]. Also, the results of the study seem to suggest that motivation is indeed a multi-dimensional construct as proposed by the Self-Determination Theory.

The second limitation is related to availability of strong theories supported by empirical studies about the relationship between different dimensions of motivation as proposed by

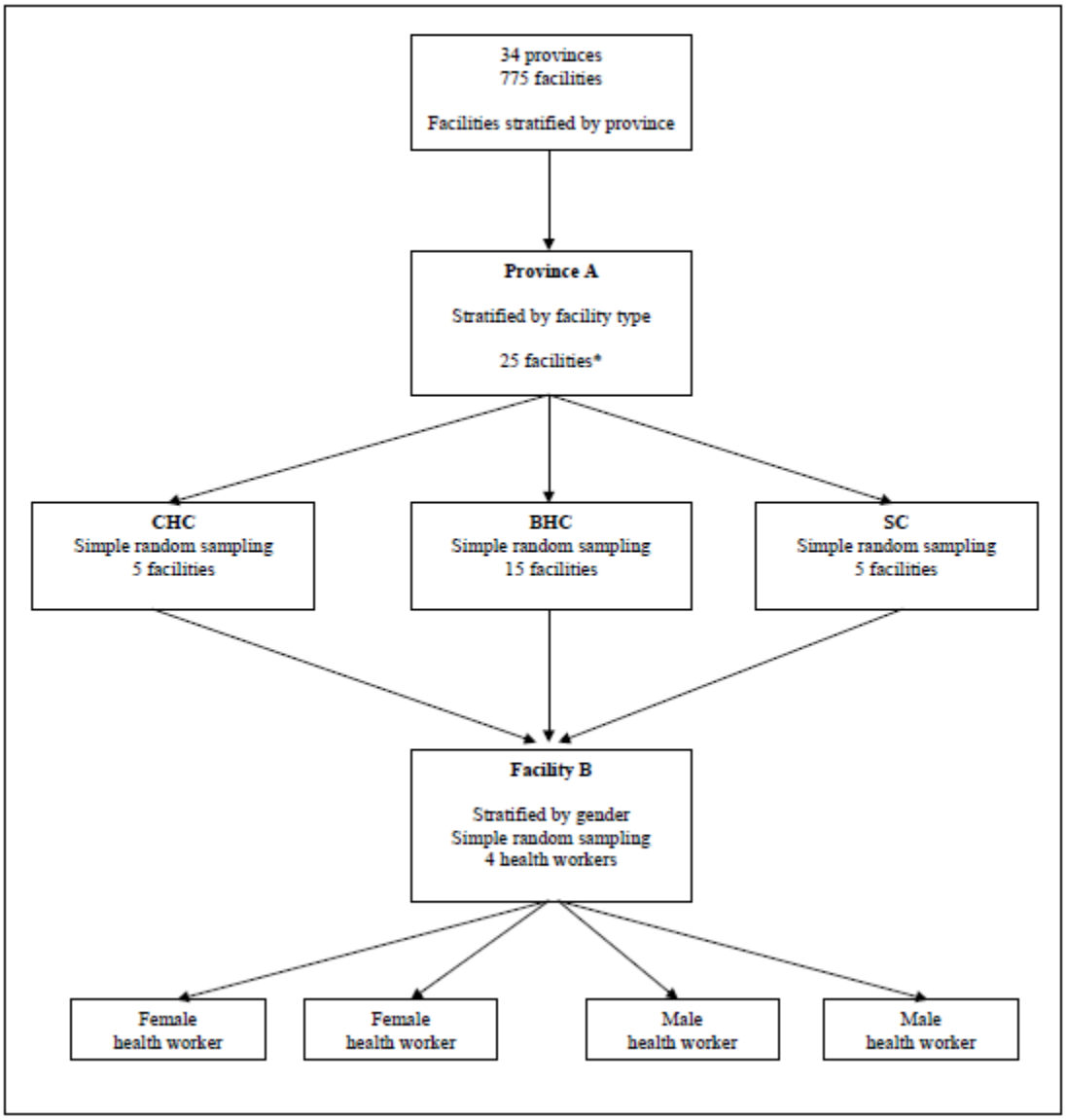
SDT and validation variables available in this study. One of the ways of demonstrating construct validity is to illustrate its relation to other constructs and behaviors. However, these constructs and behaviors must have strong theoretically plausible and empirically justified relationships to the construct of interest. As discussed earlier, there appear to be few specific studies on introjected and external regulation factors and their relationships to constructs such as staff voice or pay inequity. This limits the potential for construct validation process of the scale and the results of this study.

The third limitation is that responses obtained through the survey are likely to have been influenced by social desirability bias. Although the questionnaire forms did not contain any names, and while confidentiality of responses was emphasized, it is likely that health workers were less willing to admit to negative feelings or choose responses that would make them appear less concerned with helping their communities than with their own financial well-being. Although no formal check was made, an indication of the presence of this bias could be found using items of the Satisfaction scale, where items on satisfaction with superiors or colleagues had very high endorsement rate. This could partly explain the high frequency of endorsement on items such as "I work in this job because I have a chance to help other people through my work" or "I work in this job to gain God's grace".

Overall, however, as discussed above, the work motivation scale has performed well. Moreover, this is the first study on work motivation based on a large probability-based sample covering the entire country. Also, it includes a range of clinical health workers,

from doctors and nurses to midwives and vaccinators. In addition, it uses multiple samples to test the factorial structure, validity and reliability of the scale. Lastly, this is the first work motivation scale that has been validated in Dari and Pashto languages in a highly traditional society with low literacy levels. Given the acute shortage and high turnover of health workers in Afghanistan and other countries in Central Asia, there is a high level of interest in improving health worker motivation among governments and international development partners. An instrument that allows policy-makers and researchers to understand the nuanced multi-dimensional nature of motivation and to accurately measure it will hopefully contribute to the development of better informed policies based on evidence.

Figure 3.1 NHSPA 2012-2013 sampling for health worker questionnaires



Note: *In provinces with less than 25 facilities, all facilities have been surveyed. In the following RBF pilot provinces – Parwan, Takhar, Kunduz & Saripul – more than 25 facilities per province were selected to ensure inclusion of matched pair facilities surveyed at baseline in 2010-2011.

Figure 3.2 Key steps in the development and validation of the work motivation scale

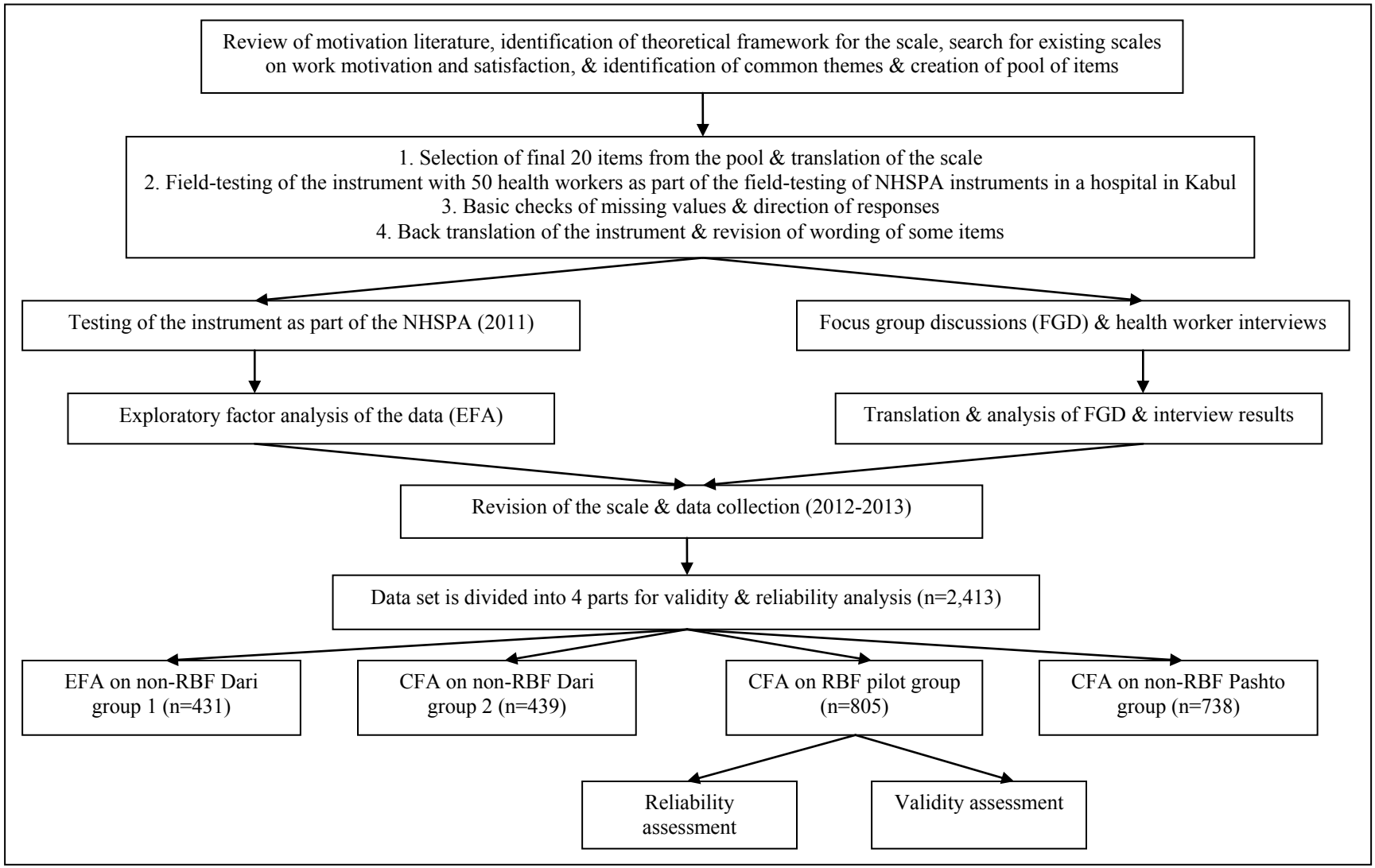


Table 3.1 Hypothesized motivation factors and their meaning

Dimension		What it means	Example
Autonomous motivation			
1	Integrated regulation^	Behavior is driven by inner self or self-identity, closest to true intrinsic motivation	A doctor who feels that being a health worker is part of her identity. She acts in ways that are consistent with her profession even outside of work by for example, helping people or treating patients outside of clinic hours.
2	Identified regulation	Behavior driven by understanding of the importance of one's duties, alignment of organizational goals with personal beliefs	A doctor who fulfills her duties because she understands the importance of providing good care to the patients, whose believes in the role of the health facility in the life of the community.
Controlled motivation			
3	Introjected regulation	Behavior driven by desire to maintain self-worth, to get approval of others, to avoid feeling of guilt	A doctor fulfilling her duties because being a doctor is a prestigious job in her community and it boosts her ego.
4	External regulation	Behavior driven by expectation of an external reward or fear of punishment	A doctor who treats her patients because she is afraid to get fired, or because she expects gifts from satisfied patients.
Lacking completely			
5	Amotivation	Absence of intent to act	A doctor who is often absent from work, avoids her duties while at work.

Note: Autonomous and controlled motivation types represented here are all part of the extrinsic motivation, which is distinguished from true intrinsic motivation. However, integrated regulation is very close to the intrinsic type of motivation, as described by the Self-Determination Theory. ^The final scale contains only four dimensions. Integrated regulation was combined with identified regulation as the result of factor analysis.

Table 3.2 Characteristics of the study population

	n	%
Survey Method		
Regular method	2,278	94.4
Community-based method	135	5.6
By type of facility		
BHC	1,367	56.7
CHC	727	30.1
Sub-Centers	319	13.2
By managing agency		
MOPH, without support	107	4.5
MOPH, with support	1,282	53.3
NGO only	1,018	42.3
By language		
Pashto	777	32.2
Dari	1,636	67.8
Gender		
Female	862	35.8
Male	1,547	64.2
Health worker type		
Doctor	336	13.9
Nurse/ Assistant doctor	580	24.0
Midwife	282	11.7
Community midwife	301	12.5
Vaccinator	690	28.6
CHW Supervisor	224	9.3
Local resident	2,011	87.7
Duration of work in this facility		
One year or less	643	28.1
2-3 yrs	714	31.2
4-5 yrs	326	14.2
6-10 yrs	467	20.4
More than 10 yrs	142	6.2
Intention to stay		
3 months or less	354	15.6
4-12 months	270	11.9
2-3 years	373	16.5
4-5 years	360	15.9
More than 5 years	906	40.0
Supervisor visit within the past 30 days	1,940	80.4
Employee assessment in the past 12 months	1,893	82.6
Of which received feedback	1,527	81.6
Salary payment up-to-date	1,578	65.5
Salary increase in the past year	881	38.4
Received performance-based payments	401	17.6
Have a second job	1,933	80.1
Technical training outside of the facility		
Number of days in the past 12 months (Mean, SD)	5.9	7.0
Technical training within the facility		
Number of days in the past 12 months (Mean, SD)	1.5	5.8
Validation group		
Non-RBF Pashto group	738	30.6
Non-RBF Dari group 1	431	17.9
Non-RBF Dari group 2	439	18.2
RBF pilot group	805	33.4

Note: Overall sample size is 2,413 observations.

Table 3.3 Eigenvalues and proportion of variance explained

Factor	Eigenvalue	Variance expl	Cumul var expl
Factor 1	5.25	0.25	0.25
Factor 2	2.88	0.14	0.39
Factor 3	1.69	0.08	0.47
Factor 4	1.19	0.06	0.52
Factor 5	1.08	0.05	0.58
Factor 6	0.92	0.04	0.62
Factor 7	0.84	0.04	0.66
Factor 8	0.82	0.04	0.70
Factor 9	0.75	0.04	0.73
Factor 10	0.72	0.03	0.77

Note: Overall sample size is 431 observations. Eigenvalues are shown only for the first ten factors.

Table 3.4 Model fit statistics, EFA of 21-item work motivation scale

Model	Chi-sq	df	CFI	TLI	RMSEA	RMSEA C.I.	SRMR
One factor	824	189	0.67	0.64	0.09	[0.082, 0.095]	0.12
Two factor	364	169	0.90	0.88	0.05	[0.044, 0.059]	0.07
Three factor	241	150	0.95	0.93	0.04	[0.028 0.046]	0.05
Four factor	176	132	0.98	0.96	0.03	[0.016 0.038]	0.04
Five factor	124	115	1.00	0.99	0.01	[0.000 0.028]	0.04

Table 3.5 Factor loadings and residual variances for a four-factor 21-item model with Geomin rotation

Item		F1 Loading	F2 Loading	F3 Loading	F4 Loading	Residual variance
intr1	I work in this job because I have a chance to help other people through my work	0.67				0.45
intr2	I work in this facility because it plays an important role in the community	0.31		0.38		0.65
intr3	I feel I should personally take the credit or blame for the results of my work on this job	0.55				0.61
intr4	I work in this job because it allows me to decide how my work is organized	0.37	0.46			0.57
intr5	I work in this job because it allows me to use my skills	0.35	0.67			0.38
intr6	I work in this job because I can accomplish something worthwhile in this job	0.46	0.39			0.57
intr7	I feel a very high degree of personal responsibility for the work I do on this job	0.52	0.34			0.58
intrj1	I work here because it makes me feel important			0.80		0.38
intrj2	I do this job because my family would be disappointed if I quit			0.42		0.73
intrj3	I do this job because it gives me respect in the community			0.77	-0.36	0.35
intrj4	I work in this job to gain God's grace	0.69		0.29		0.47
extrg1	I only work here to get paid	-0.28		0.30		0.77
extrg2	I work here because of opportunities for promotion		0.47	0.21		0.61
extrg3	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)		0.71	-0.34		0.59
extrg4	I work here because it is located in a safe area		0.50		-0.44	0.52
extrg5	I work here because of good benefits I receive (Note: all benefits – housing, transportation, anything else you receive – think overall)	-0.31	0.43			0.70
extrg6	I work here because it provides long term security for me		0.50			0.67
amot1	I frequently think of quitting this job	-0.33		0.19	0.46	0.58
amot2	I sometimes feel my work here is meaningless	-0.45			0.46	0.56
amot3	I don't care much about the quality of work here	-0.62				0.58
amot4	I work here because I have no other choice		0.22	0.34		0.73

Note: Only loadings significant at 5% level or minimum value of 0.30 are presented here.

Table 3.6 Factor structure for a four-factor 21-item model with Geomin rotation

Item		F1	F2	F3	F4
intr1	I work in this job because I have a chance to help other people through my work	0.70	0.34	0.18	-0.02
intr2	I work in this facility because it plays an important role in the community	0.35	0.39	0.45	-0.01
intr3	I feel I should personally take the credit or blame for the results of my work on this job	0.55	0.23	0.25	0.14
intr4	I work in this job because it allows me to decide how my work is organized	0.44	0.54	0.27	0.01
intr5	I work in this job because it allows me to use my skills	0.46	0.71	0.23	-0.11
intr6	I work in this job because I can accomplish something worthwhile in this job	0.53	0.47	0.17	-0.10
intr7	I feel a very high degree of personal responsibility for the work I do on this job	0.57	0.39	0.10	-0.02
intrj1	I work here because it makes me feel important	0.10	0.33	0.78	0.11
intrj2	I do this job because my family would be disappointed if I quit	-0.07	0.26	0.49	0.24
intrj3	I do this job because it gives me respect in the community	0.03	0.38	0.72	-0.23
intrj4	I work in this job to gain God's grace	0.68	0.13	0.25	0.01
extrg1	I only work here to get paid	-0.27	0.21	0.37	0.18
extrg2	I work here because of opportunities for promotion	0.24	0.58	0.41	0.01
extrg3	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)	0.10	0.56	-0.04	-0.11
extrg4	I work here because it is located in a safe area	-0.05	0.53	0.22	-0.43
extrg5	I work here because of good benefits I receive (Note: all benefits – housing, transportation, anything else you receive – think overall)	-0.24	0.43	0.31	0.07
extrg6	I work here because it provides long term security for me	0.14	0.57	0.34	-0.02
amot1	I frequently think of quitting this job	-0.38	-0.06	0.24	0.52
amot2	I sometimes feel my work here is meaningless	-0.48	-0.03	0.11	0.50
amot3	I don't care much about the quality of work here	-0.63	-0.07	0.04	0.21
amot4	I work here because I have no other choice	0.00	0.36	0.46	0.21

Table 3.7 Model fit statistics, CFA of Models A and B

Model	Chi-sq	df	CFI	TLI	RMSEA	RMSEA C.I.
Model A	267	98	0.87	0.84	0.06	[0.05, 0.07]
Model B	193	84	0.91	0.89	0.05	[0.04, 0.06]

Note: Model A is a 16-item scale. Model B is a 15-item scale. The overall sample size is 439 observations.

Table 3.8 Factor loadings of a four-factor 15-item model (Model B)

		Loading	S.E.	R-Sq
Identified regulation F1				
intr1	I work in this job because I have a chance to help other people through my work	0.68	0.06	0.46
intr3	I feel I should personally take the credit or blame for the results of my work on this job	0.58	0.05	0.34
intr7	I feel a very high degree of personal responsibility for the work I do on this job	0.70	0.07	0.49
intrj4	I work in this job to gain God's grace	0.67	0.07	0.44
External regulation F2				
extrg2	I work here because of opportunities for promotion	0.58	0.05	0.33
extrg3	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)	0.37	0.06	0.14
extrg4	I work here because it is located in a safe area	0.32	0.06	0.10
extrg6	I work here because it provides long term security for me	0.57	0.05	0.33
intr5	I work in this job because it allows me to use my skills	0.69	0.05	0.48
Introjected regulation F3				
intrj1	I work here because it makes me feel important	0.72	0.04	0.51
intrj2	I do this job because my family would be disappointed if I quit	0.57	0.05	0.32
intrj3	I do this job because it gives me respect in the community	0.76	0.05	0.58
Amotivation F4				
amot1	I frequently think of quitting this job	0.54	0.05	0.29
amot2	I sometimes feel my work here is meaningless	0.69	0.06	0.48
amot3	I don't care much about the quality of work here	0.71	0.05	0.51

Note: All factor loadings are significant at 5% level.

Table 3.9 Factor correlations (Model B)

	F1	F2	F3	F4
F1				
F2	0.61**			
F3	0.49**	0.68**		
F4	-0.69**	-0.50**	0.02	

Note: **significant at 5% level.

Table 3.10 Model fit statistics, CFA of 15-item work motivation scale, RBF sample

Model	Chi-sq	df	CFI	TLI	RMSEA	RMSEA C.I.
Model B	253	84	0.93	0.91	0.05	[0.04, 0.06]

Note: The overall sample size is 805 observations. However, one observation is missing values on all items, reducing the sample size to 804 observations.

Table 3.11 Factor loadings of a four-factor 15-item model, RBF sample

		Loading	S.E.	R-Sq
Identified regulation F1				
intr1	I work in this job because I have a chance to help other people through my work	0.66	0.04	0.44
intr3	I feel I should personally take the credit or blame for the results of my work on this job	0.58	0.04	0.33
intr7	I feel a very high degree of personal responsibility for the work I do on this job	0.68	0.05	0.46
intrj4	I work in this job to gain God's grace	0.69	0.06	0.47
External regulation F2				
extrg2	I work here because of opportunities for promotion	0.49	0.04	0.45
extrg3	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)	0.67	0.03	0.24
extrg4	I work here because it is located in a safe area	0.47	0.03	0.22
extrg6	I work here because it provides long term security for me	0.54	0.04	0.29
intr5	I work in this job because it allows me to use my skills	0.67	0.03	0.45
Introjected regulation F3				
intrj1	I work here because it makes me feel important	0.73	0.04	0.49
intrj2	I do this job because my family would be disappointed if I quit	0.70	0.04	0.19
intrj3	I do this job because it gives me respect in the community	0.44	0.04	0.54
Amotivation F4				
amot1	I frequently think of quitting this job	0.63	0.04	0.39
amot2	I sometimes feel my work here is meaningless	0.81	0.03	0.66
amot3	I don't care much about the quality of work here	0.70	0.04	0.48

Note: All factor loadings are significant at 5% level.

Table 3.12 Factor correlations, RBF sample

	F1	F2	F3	F4
F1				
F2	0.68			
F3	0.31	0.72		
F4	-0.60	-0.27	0.28	

Note: All correlations are significant at 5% level.

Table 3.13 Inter-item and item-total correlations and ordinal alpha

Identified F1	**intr1**	**intr3**	**intr7**	**intrj4**		**alpha**
intr1						
intr3	0.36					
intr7	0.40	0.41				
intrj4	0.39	0.33	0.60			
f1t	**0.36**	**0.33**	**0.45**	**0.42**		**0.74**
External F2	**extrg2**	**extrg3**	**extrg4**	**extrg6**	**intr5**	**alpha**
extrg2						
extrg3	0.30					
extrg4	0.31	0.31				
extrg6	0.40	0.27	0.25			
intr5	0.45	0.42	0.20	0.29		
f2t	**0.46**	**0.42**	**0.35**	**0.39**	**0.43**	**0.70**
Introjected F3	**intrj1**	**intrj2**	**intrj3**			**alpha**
intrj1						
intrj2	0.27					
intrj3	0.52	0.34				
f3t	**0.43**	**0.32**	**0.47**			**0.64**
Amotivation F4	**amot1**	**amot2**	**amot3**			**alpha**
amot1						
amot2	0.52					
amot3	0.43	0.56				
f4t	**0.45**	**0.55**	**0.49**			**0.75**

Note: Item-total correlations are estimated ommitting the item in question from the total.

Table 3.14 Construct validity variables and rationale

Construct validity variable	Factor	Expected association direction and rationale
Intention to quit	Amotivation	Positive: Amotivation as lack of intention to act is likely to be associated with desire to quit. Plus, literature on low motivation and staff turnover (e.g. Shields & Ward, 2001; Yildiz et al, 2009)
Performance assessment & feedback	Identified regulation	Positive: Meaningful performance feedback promotes a sence of competence and autonomy (Gagne & Deci, 2005; Deci et al, 1989)
Staff voice	Identified regulation	Positive: Staff feels that their managers care about their point of view and acknowledge their perspectives. This increases the sense of competence and autonomy among employees (Deci et al, 1989; Greco & Eisenberg, 1993)
	Introjected regulation	Positive: SDT is not clear about this association, but to the extent that staff voice as ability to express one's opinions increases one's self-esteem and promotes the feeling of self-worth, it is reasonable to expect a positive association.
Lack of fairness of the supervisor	Identified regulation	Negative: Fairness is an important predictor of autonomous work motivation (Gagne & Forest, 2011). Theory on organizational justice predicts that consistent treatment, bias-free judgements and clear standards influenced organizational citizenship behaviors and were associated with higher commitment to the organization (Mitchell, 1997).
	Amotivation	Positive: Theory on organizational justice predicts that feelings of unfairness & inequity were associated with withdrawal, withholding of effort and reduced persistence (Mitchell, 1997).
Equity of rewards	Identified regulation	Positive: While external rewards are expected to have negative association with identified motivation, organizational justice, including distributive justice, is expected to have a positive association with this type of motivation. It is reasonable to expect a positive association between the two variables if one views equity of rewards as one of the elements of organizational justice (Gagne & Forest, 2011).
	External regulation	Positive: External rewards are expected to have positive association with external regulation (Gagne & Deci, 2005). SDT does not seem to focus explicitly on organizational justice and external regulation, but it seems reasonable to hypothesize that if one is motivated by extrinsic rewards such as pay, their motivation is influenced by whether they believe they are being rewarded fairly for their work. Also, organizational justice theory that views motivation as unidimensional predicts association between pay inequity and low motivation (Mitchell, 1997).
	Introjected regulation	Positive: SDT does not seem to focus explicitly on organizational justice and introjected regulation, but pay can be viewed as informational feedback. It seems reasonable then to hypothesize that when employees are motivated because they want to maintain feelings of self-worth, if they perceive that they are being under-rewarded, they feel less motivated.

Table 3.15 Comparison of factor means between those intending to quit and those intending to stay

	Estimate	
Model fit statistics		
AIC	25682	
BIC	25773	
Entropy	1.0	
Means	Estimate	S.E
Identified F1	0.04	0.14
External F2	0.29	0.16
Introjected F3	0.27	0.16
Amotivation F4	-0.44**	0.15

Note: The overall sample size is 795 observations. BIC is sample-size adjusted.
** significant at 5% level.

Table 3.16 Staff voice and equity as predictors of motivation factors

	Express opinion		Unfair supervisor		Rewarded fairly	
	Estimate	d.f./CI	Estimate	d.f./CI	Estimate	d.f./CI
Model fit statistics						
Chi-square	273	95	271	95	278	95
CFI	0.92		0.93		0.93	
TLI	0.90		0.91		0.91	
RMSEA	0.05	[0.04, 0.06]	0.05	[0.04, 0.06]	0.05	[0.04, 0.06]
	Beta	S.E.	Beta	S.E.	Beta	S.E.
Coefficients						
Identified F1	0.23**	0.04	-0.21**	0.03	0.01	0.04
External F2	0.33**	0.03	-0.05	0.02	0.33**	0.03
Introjected F3	0.14**	0.04	0.07	0.03	0.21**	0.03
Amotivation F4	-0.09**	0.03	0.47**	0.03	0.02	0.03

Note: The overall sample size is 803 observations. Coefficients are standardized. ** significant at 5% level.

Table 3.17 General motivation item as a measure of construct validity of the scale

	Estimate	d.f./CI
Model fit statistics		
Chi-square	287	95
CFI	0.93	
TLI	0.91	
RMSEA	0.05	[0.04, 0.06]
	Correlation	Validity coefficient
Coefficients		
Identified F1	0.68**	0.59
External F2	0.44**	0.36
Introjected F3	0.13**	0.11
Amotivation F4	-0.35**	-0.30

Note: The overall sample size is 804 observations. ** significant at 5% level.

Table 3.18 Measurement equivalency test of Dari and Pashto language scales using CFA

Model		Chi-sq	df	CFI	TLI	RMSEA	RMSEA C.I.
1	Invariance of factorial structure	495	168	0.92	0.90	0.05	[0.05, 0.06]
2	Invariance of factor loadings	522	179	0.92	0.91	0.05	[0.05, 0.06]
3	Invariance of thresholds & factor loadings	684	205	0.89	0.89	0.06	[0.05, 0.06]

Note: Overall sample size for Dari is 804 observations, and for Pashto 738 observations.

4 Chapter 4: Impact of performance-based payments on health worker motivation in Afghanistan (Paper 2)

Abstract

Background: Despite some progress in improving health outcomes over the past decade, Afghanistan still has one of the highest maternal mortality ratios in the world. It is also struggling with extremely high rates of infant and child mortality. Access to health services depends vitally on availability of health workers. To attract and retain qualified health workers, it is necessary to improve health worker motivation. Moreover, motivation also appears to be an important determinant of health worker performance. This study examines the effect of performance-based payments on health worker motivation in Afghanistan. Motivation is defined as a multi-dimensional construct consisting of the following four factors: identified regulation, external regulation, introjected regulation and amotivation.

Method: In 11 provinces of Afghanistan, 442 facilities were randomly assigned to an intervention or a comparison arm. Because of the study design, masking of the participants was not feasible. The study is based on a cross-sectional survey of 805 health workers from 256 facilities conducted 23 months after the start of the intervention. The outcome, health worker motivation, was measured using a 15-item four-factor scale. A robust variance estimator was used to adjust for the correlation of the data within health facilities. A latent variable modeling framework was used to estimate the effect of performance-based payments on the four motivation factors. The effect of treatment was

estimated using the intention-to-treat analysis. However, given the high rate of non-compliance, a complier-average causal effect approach was also examined.

Results: No statistically significant changes were observed in health worker motivation in the intervention group as compared to the control group (p-value>0.05). Financial incentives in the form of performance-based payments did not seem to have an effect on health worker motivation factors. The estimated complier-average causal effects for identified (γ=-2.63, p-value<0.05), external (γ=-1.40, p-value<0.05), and introjected (γ=-1.37, p-value<0.05) motivation factors were negative. It appeared that PBP were negatively associated with motivation.

Conclusion: The negative effect of PBP on identified motivation supports the overall hypothesis of the study based on the Self-Determination Theory of motivation. However, the negative effect on external motivation was surprising. Yet, there are several design issues of the PBP program in Afghanistan that may explain these unexpected results. First of all, there were heightened expectations that were not met due to the relatively small size of incentive payments in proportion to total salaries and payment delays. Secondly, it is unclear whether health workers were able to distinguish these payments from their regular salaries and other financial benefits. Thirdly, a large portion of facilities distributed PBP earned by the facility in proportion to existing salaries, which meant that the most senior staff or staff with highest salary levels received the largest portion of PBP and not the staff most directly responsible for the achievement of the performance indicators, such as midwives and vaccinators. Fourthly, PBP were not coupled with

improvements in working conditions at the facility level. The study thus demonstrates the importance of design and implementation details of PBP programs and identifies specific aspects of the program where changes could be made.

4.1 Introduction

Health system performance depends vitally on availability of health workers, and the growing crisis in human resources will have a high negative impact on health outcomes [135]. According to the World Bank Development Report*, Making Services Work for Poor People*, weak incentives for performance is one of the key reasons why health services, among others, fail the poor. Performance-based payments (PBP) are thought to be one of the mechanisms of increasing provider accountability to their patients or rather clients, and thus, improving quality of health care services for them [136]. PBP or supply-side incentive programs under the Results-Based Financing (RBF) projects in health have been gaining popularity in many developed and developing countries over the past few years [8, 43]. There are 30 countries where the RBF in health projects were recently completed, are currently implemented or under the preparation [95]. While RBF projects include demand- and supply-side incentives, this provides some indication of the large amount of resources that are spent on improving health system performance through changes in provider incentives.

In the summer of 2010, Afghanistan launched an RBF project in purposefully selected 11 of its 34 provinces (for details on RBF design see Chapter 1 and [68]). The project aimed "to impact MDGs 4 and 5 by improving coverage of maternal and child health services within the existing health system and without creating unnecessary parallel processes" [10]. Its primary objective was "to increase key maternal and child health outputs, to improve quality of health services and to ensure higher patient satisfaction with health service delivery" [10]. It was hypothesized that improved financial incentives in the form

of pay-for-performance would improve health worker motivation and satisfaction, which in turn would lead to better performance.

One of the key assumptions of these programs is that providing financial incentives based on performance will improve motivation of health providers and consequently, their retention and performance. However, a review of the literature by Eldridge and Palmer [41] did not find strong evidence base for the success of PBP programs in developing countries. This view is supported by a more recent review by Eijkenaar et al [137]. The evidence base for the effectiveness of PBP programs has even more gaps when it comes to their impact on health worker motivation. The jury is still out on the role of external rewards, specifically financial incentives, on work motivation [19, 43, 54, 138]. The aim of this paper is therefore to assess health worker motivation in Afghanistan in health facilities allocated to performance-based payments as compared to health facilities allocated to control.

Work motivation which is willingness to exert and maintain an effort to achieve the organization's goals is defined as a multi-dimensional construct consisting of four factors: (1) identified regulation, (2) external regulation, (3) introjected regulation, and (4) amotivation. Identified regulation describes behavior in which people act because the activity is congruent with their personal goals and beliefs. Gagne and Deci [19] give an example of a nurse who performs uninteresting and not necessarily pleasant tasks such as bathing patients because she values her patients' health and understands the importance of these activities for their well-being. Identified regulation is considered to be an

autonomous type of motivation, while external regulation is the controlled type. External regulation describes behavior that is driven only by an expectation of a reward or fear of negative consequences. Introjected regulation is like a bridge between controlled and autonomous motivation. The regulation comes from within but it is controlled by external forces. Examples of this type of regulation include "contingent self-esteem, which pressures people to behave in order to feel worthy, and ego involvement, which pressures people to behave in order to buttress their fragile egos" [19]. Introjected regulation lies between the identified and external regulation. Amotivation is defined as the absence of intent to act in a particular way [22]. As described in detail in Chapter 1, according to SDT, extrinsic contingent rewards undermine the feeling of autonomy and have a negative effect on more autonomous types of motivation such as identified regulation. It is hypothesized that external regulation has a positive association with extrinsic rewards.

4.1.1 Research questions

The study aims to answer the following research questions:

1. What is the effect of PBP program on different types of motivation, i.e. the average effect comparing health workers in the treatment arm to those in the control arm?
    a. Are identified and introjected motivation types lower among health staff working in treatment facilities as compared to control?
    b. Is there a positive effect of PBP program on externally regulated motivation in treatment group as compared to control?

c. Does PBP program lower amotivation in treatment group as compared to control?

2. What is the efficacy of the performance-based payments, i.e. the effect of payments on health workers who actually received the payments?

4.2 Methods

4.2.1 Data source

The data for this study was derived from the cross-sectional health facility survey for the annual National Health Services Performance Assessment (NHSPA) implemented in Afghanistan over the period of nine months between 2012 and 2013. The survey was based on a multi-stage stratified probability sampling approach where primary sampling unit was a health facility (Figure 4.1):

- Stratification by province,
- Stratification by type of health facility within each province,
- Selection of facilities based on simple random sampling,
- Stratification of health workers within facilities by gender,
- Selection of individual health workers.

Individual health workers were secondary sampling units. Within each facility four health workers classified as clinical staff (doctors, nurses, assistant doctors, midwives, community midwives, and vaccinators) present at the day of the survey were randomly selected for the survey. If fewer than four clinical staff were present at the facility, then other facility staff members, including community health supervisors, pharmacists, cleaners and others, were interviewed for the survey. However, all non-clinical workers

with exception of community health supervisors were excluded from this analysis. This resulted in a sample of 256 health facilities with 805 health workers, or 3.1 health workers per facility. The unit of analysis is individual health worker.

4.2.2 Variables construction

Outcome variables

Health worker motivation

This is a latent variable measured through a self-administered 15-item scale. As factor indicators were measured on a four-point Likert scale, the outcome is considered to be an ordinal variable, although the unobserved variable underlying each indicator is continuous. There are four factors in the scale: (1) identified motivation consisting of four items, (2) external regulation consisting of five items, (3) introjected regulation consisting of three items, and (4) amotivation consisting of three items. The scale is based on Self-Determination Theory. There is no summary score for the scale or its four sub-scales. The details of the work motivation scale construction, its validity and reliability estimates are described in Chapter 3. The variable is measured at an individual health worker level.

Predictor variables

Treatment

This is the main predictor variable that is measured at a cluster (facility) level. It is a binary variable based on a random allocation of facilities to either treatment or control at the start of the RBF program. Facilities were assigned to either treatment or control based on the information provided by MOPH.

Compliance

Respondents were provided with a list of benefits and asked to check all benefits that applied to them ("Do you currently receive any of the following benefits?"). The list of benefits included usual benefits such as free housing, transportation allowance, meals at work, and seniority payments, but it also included pay-for-performance. Next to the question explanatory note was added on the meaning of pay-for-performance: "Makafat Naqdi is when someone gets extra money in addition to their regular salary because they have completed certain tasks." This explanatory note was developed by Afghan staff involved in design and pilot of the questionnaire. Those who were in treatment group and marked pay-for-performance among the list of their benefits were considered to be "compliers", while those who did not mark it were classified as "non-compliers". As this is an observed compliance status and it could not be observed in those assigned to control group, it has missing values for all those who were assigned to control group.

Facility type

Different types of facilities differ in staffing norms, infrastructure, and services that they are mandated to perform and size of the catchment population [66]. Thus, this is a good proxy variable for the size of the facility and level of infrastructure. As described in section 1.4 on RBF in Afghanistan, SC is the most basic type of health facility included in the program. Each SC serves about 3,000 – 7,000 people. It should have 1 male nurse and 1 female community midwife [66]. BHC has a catchment area of 15,000 – 30,000 people and has the following staffing norms: 1 community health supervisor, 2

vaccinators, 1 male nurse, 1 community midwife, 1 general physician, preferably female [66]. CHC has a catchment area of 30,000 – 60,000 people and has the following staffing norms: 1 community health supervisor, 2 vaccinators, 1 male nurse, 1 female nurse, 2 community midwives, 1 male general physician, 1 female general physician, 1 laboratory technician, 1 pharmacy technician, and one (male or female) psychosocial counselor [66]. CHC provides specialized outpatient care and basic inpatient care. Based on the original categorical variable, dummy variables were created for the analysis where SC was used as a reference group.

Management type

There are three main types of management of health facilities in Afghanistan: (1) MOPH only, (2) MOPH with institutional support, and (3) NGOs only. The latter is further sub-divided by type of contracting which is not one of the study variables here. Facilities run by MOPH only, MOPH with support, and NGOs only differ in management structure, source of funding, and its levels [39]. While there are varying degrees of managerial autonomy within the third group, facilities in this group have the highest level of independence when compared to the other two groups [64]. Thus, it is an important variable containing information on many aspects of facility management – such as funding level, organizational structure and culture, ability to take timely decisions on things like staffing, system of rewards, infrastructure – that were not otherwise measured in this survey.

Equipment functionality index

The equipment functionality index consists of 20 items for BHCs and SCs and 23 items for CHCs. It includes items such as children's scales, sterilizer, stethoscope and others that are mandated by the BPHS Guidelines[66]. The composition of the index was based on the methodology provided in the BSC used in Afghanistan for evaluation of health system performance since 2004 [40]. An item was rated as 1 if it was available and was in working condition, and 0 otherwise. Following this, the scores for each item were added together and divided by the total number of items (20 for BHCs and SCs and 23 for CHCs). The variable was standardized as a z-score based on the overall population or grand mean. Due to its highly negatively skewed distribution it was transformed (original^3) prior to testing of differences in means between the two treatment groups [103]. The index was measured at the facility level.

Pharmaceuticals & vaccine availability index

This index consists of 31 items that are considered to be the essential medicines and vaccine that must be available at all BPHS facilities. Similar to equipment functionality index it was based on the methodology provided in the BSC revised by the national technical working group led by MOPH [40]. The item was rated as 1 if it was continuously available for the past 30 days, and 0 otherwise. The variable was standardized as a z-score based on the overall population or grand mean. Due to its highly negatively skewed distribution it was transformed (original^3) prior to testing of differences in means between the two treatment groups [103]. The index was measured at the facility level.

Infrastructure index

This index consists of nine items that describe presence and condition of basic infrastructure in the facility, such as heat, electricity, water source, windows etc. The original index used in the BSC [40] was composed of ten items. However, one of the items on presence and condition of exterior walls had very high proportion of missing values, which resulted in 14.5% of missing values for the index as a whole. Given this high missingness and that this particular item seemed to indicate less about the quality of services at facilities than items such as reliable electricity and water supply, or presence of interior walls and windows, it was removed from the index [4]. The item was rated as 1 if was present and few or no repairs were needed, and 0 otherwise. The variable was standardized as a z-score based on the overall population or grand mean. It was approximately normally distributed.

Pay-for-performance received by the facility

Respondents were asked whether or not their facility received pay-for-performance "Has your facility received performance-based payments?" Unlike the treatment variable which was based on the MOPH information, this variable was reported by health workers themselves. Thus, it reflected the awareness of health workers regarding PBP in their facility. This question also differs from the one on benefits described above under Compliance variable description. It asks whether facility, not the respondent personally, has received pay-for-performance.

Health worker type and gender

The health worker questionnaire containing the Work motivation scale also contained questions on basic health worker characteristics. Gender is a binary variable, where 1 means male. Health worker type is categorical variable. Health workers were asked to mark one of the following categories: doctor, nurse, midwife, community midwife, CHW supervisor, vaccinator, or other. The last category included administrator, laboratory technicians, pharmacists, guards, cleaners, and other non-clinical staff. These were excluded from the current study. The categorical variable was transformed into dummy variables where doctors were a reference group.

Other variables used for health worker characteristics

Duration of work in this facility was originally continuous variable that was made into a categorical variable with the following five categories: (1) one year or less, (2) two to three years, (3) four to five years, (4) six to 10 years, and (5) more than ten years. The original continuous variable had a very big range of values (0 days to 30 years), many gaps in them, and was highly skewed. For the purposes of describing the sample and checking comparability of two treatment arms or compliance groups, it seemed more useful to redefine this variable as categorical.

Supervisor visit within the past 30 days is a binary variable. There were no follow-up questions on the purpose or content of the visit. Formal employee assessment in the past 12 months is also a binary variable with a follow-up question on whether there was a feedback from the assessment. Only the question on provision of assessment was included in this study (see Chapter 3 on discussion of feedback and assessment variables

as related to motivation). Salary payment up-to-date is a binary variable that has been one of the key indicators in the BSC for assessing health system performance in Afghanistan [40]. Salary increase in the past year is also a binary variable. Technical training outside of the facility is a measured as the number of days and is continuous. It had high positive skew and was transformed through the square root that is often used with positively skewed variables [103].

Given the methodology used in the study as described in the next section, there was a hard constraint on the number of predictors that could be used in the model. Most of the variables described above were used only for examining the success of randomization by comparing the two treatment arms. They were also used to examine differences between compliers and non-compliers, but only a fraction of them was selected for the actual model. Selection was based on the results of exploratory data analysis and theoretical understanding of the potentially most important covariates in the relationship between pay-for-performance and motivation.

4.3 Analysis

4.3.1 Missing data

Overall, there were very few missing values for the variables used in the study. There were no variables that had more than 40 missing observations, which was still only 5% of the sample. Moreover, proportion of missing values was less than 1.5% (11 out of 805 observations) in all but three variables. The three variables with the highest proportion of missingness (4% – 5%) were the three facility level quality indices. Each of these indices

consists of a number of items, ranging from 9 items in Infrastructure Index to 31 in Pharmaceuticals and vaccines availability index (See section on variables construction for details). Thus, presence of even few missing values in one of the items results in high proportion of missingness for the index overall. While unconditional mean imputation can result in reduction of variance and inferences can be distorted by bias and overstated precision as discussed by Little [139], for scales with multiple items use of the item mean and individual mean are acceptable methods [140, 141]. Studies on missing values in multi-item scales recommend imputing the individual's mean as robust and simple method of imputation and as an alternative to multiple imputation method [140, 141]. However, these studies examine cases with 10% - 30% missing values. Given the low proportion of missingness in the current case ($\leq$ 5%), we used mean values for treatment and control groups respectively to impute the missing values for these three indices. For all other variables, given the small proportion of missingness, no imputations were performed.

Self-reported pay-for performance or Compliance variable (Do you currently receive any of the following benefits?") had eight missing values originally (1.0%). However, five of these were in health facilities that did not receive pay-for-performance according to the same respondents and in addition, were part of the control group. Thus, as it was assumed that health workers could not receive performance-based payments in facilities where none existed, these were corrected from missing to zero ("No"). After this correction, only three out of 805 observations (<0.5%) had a missing value for this variable.

4.3.2 Analysis approach

First, all study variables were checked for consistency. The variable on whether a health worker personally received pay-for-performance (Compliance) was checked for consistency with the variable on whether a facility received pay-for-performance, also reported by health workers in the same questionnaire. It was also cross-checked with whether facility was assigned to treatment or control. Only ten observations assigned to control reported receiving pay-for-performance. These were not concentrated by facility, i.e. one out of four health workers surveyed in a given facility reported receiving pay-for-performance. Based on this, it appeared that these ten observations were likely to be errors in recording and were corrected to zero ("No").

Pay-for-performance received by facility as reported by health workers (as opposed to treatment variables based on MOPH data) was largely consistent with treatment variable. Only 7% of those in the control group (n=375) reported that their facility received pay-for-performance. A portion of these could be errors in recording or misunderstanding on the part of these twenty-seven health workers. However, two thirds of these health workers were concentrated by facility: They came from seven facilities where more than one health worker reported that their facility received pay-for-performance. These were likely to come from facilities that may have pay-for-performance outside of RBF as MOPH, according to the National Salary Policy [142], encourages implementing agencies managing facilities to experiment with performance-based payments. Thus, these were not corrected and allowed to differ from the treatment variable.

Gender values were corrected based on health worker position – all midwives and community midwives were coded as females because according to BPHS Guidelines these positions must be filled by female staff [66].

All categorical variables were examined using frequency distribution tables and bar graphs. Distribution of each item of the work motivation scale was examined separately. Simple summary statistics (means, medians, standard deviations and maximum and minimum values) were used to examine continuous variables. Stem-and-leaf plots, histograms and box plots were used to do visual checks and examine the symmetry, center, spread and outliers for continuous variables. Outliers were first checked using box plots and more formally with "lv" function in STATA [103]. Values above or below the inner fences are considered to be mild outliers and values above or below the outer fences are considered to be severe outliers [103]. Severe outliers were further examined and replaced where they appeared to be the result of error. In addition, dot plot showing facility mean and 95% confidence interval was constructed for each item to examine spread of facility means in relation to the grand mean (mean for the overall sample). This provided a sense of the variation in facility means.

Following the exploratory data analysis, selected characteristics of the sample were examined by treatment allocation. Means and standard deviations were estimated for continuous variables, while for binary and categorical variables estimates of proportions were used. Data were weighted to account for unequal probability of selection at the health worker level as it is recommended for descriptive analysis of population

characteristics [143]. Test statistics were adjusted for multi-stage stratified cluster sampling using Taylor-linearized variance estimation using the survey commands in Stata [105]. Strata with single sampling unit were centered at overall mean. These steps were done to help in identifying potential confounders by examining the sample variations. Similar analysis was performed for compliers and non-compliers as defined by the Compliance variable (see previous section). Examining key characteristics of the sample in this way was important for identifying good predictors for the compliance status used in CACE modeling later.

After these exploratory analyses, structural equation modeling approach for complex survey data was used to estimate the effect of treatment on motivation factors. According to Rabe-Hesketh, Skrondal and Pickles [144], structural equation models are generally required with latent variables as they are measured through several observed indicators (scale items) resulting in multiple responses. Given that our primary interest is the association between the latent variable (motivation) and its predictors (RBF treatment), we have two parts to our model: (i) measurement part which measures the association between the observed indicators and the latent construct, and (ii) structural part which measures the association between the latent construct and its predictors. There are three alternative approaches to measuring these two parts [145]:

1. Treat each response variable, i.e. latent variable indicator, individually. This will require fitting m regressions that describe individually the association between the predictor variable and each response variable, Y. The main weakness of this approach is that we will not be able to test our hypothesis because we will come

to different conclusions for each of the indicator but will not be able to come to infer any association between our latent variable of interest and predictors.

2. Summarize multiple responses first and then test the association between the response and the predictors. In this approach we are basically treating what is an imperfectly measured latent variable as an observed variable. This summary variable is an "imperfect" measure of our latent construct with reliability less than 1. In general, this leads to overestimating the variance of our outcome (Var $T(x)<Var(X)$ where $T(x)$ is our true outcome and X is what we are able to measure). The consequence of overestimating the variance is increased standard errors of our regression coefficients, leading to increased Type II error, where we fail to reject the null hypothesis when the alternative is true. As per Bollen [146], in multiple regressions the consequences of measurement error are unknown and can lead to coefficients that are higher, lower or the same as true coefficients.
3. Summarize and analyze simultaneously using structural equation models (SEM). The strength of SEM is that both measurement and structural parts are fit simultaneously in one step. Thus, it does not assume that the latent variable is a simple summary of observed indicators and it is known. This acknowledges the measurement error inherent in measurement of latent constructs through observed indicators.

The study used two approaches to the analysis: (a) intention-to-treat, and (b) complier average causal effect. Intention-to-treat approach is the "gold standard" in the analysis of randomized controlled trials [147, 148]. Intention-to-treat approach analyzes data based

on the initial random assignment of treatment. In this way, it tries to preserve the original intent of random allocation, where all participants have equal chance of receiving a treatment and therefore, there are no systematic differences between the groups allocated to different treatments with regard to known or unknown characteristics. If everyone allocated to a particular treatment group followed her assignment and there was no missing data, using ITT approach would give results reflecting the principle of randomization[149].

However, as it was the case with the current study, many randomized trials suffer from treatment non-compliance or intervention non-adherence. In this case, ITT approach may result in decreased power to detect the effect of the treatment. In fact, ITT will always give more attenuated effect than an approach that accounts for non-adherence [149, 150]. As Little and Yau put it, "this estimates the causal effect of treatment assignment rather than the effect of the treatment for participants who actually received it" [151]. However, the ITT provides the estimate of the effectiveness of the program, as in real life situation there will always be compliers and non-compliers and if one is to evaluate whether the program works overall, ITT is the most appropriate estimate[149]. Thus, we used ITT approach first.

Given the limitations of the ITT approach described above and the interest in effect of performance-based payments on health worker motivation beyond programmatic effects, complier average causal effect (CACE) was used for further analysis.

CACE is the method of analysis where, in simple terms, the effect of intervention is estimated based on those who were assigned to and actually received the treatment. It is best illustrated with a simple example of smoking cessation program where a group of individuals has been randomized into two groups: group one will receive a week-long training and group two will receive "treatment as usual", i.e. no special services. At the end of the training program, those who completed all sessions in group one will receive certificates. Assuming that in group one a large proportion of individuals fails to complete the training, comparing group one with group two will give attenuated results on the effect of the training on smoking cessation. It is easy to agree that those who completed the training differ from those who failed to do so. For example, they could have been more motivated to quit smoking or less addicted to it. Thus, to have a true counterfactual, one needs to identify similar types of participants in the control group. Comparing those who completed the training in the treatment arm to the same types of participants, i.e. those who would have completed the training had they been assigned to group one, gives the complier-average causal effect. It differs fundamentally from the as-treated approach, because the latter violates randomization and ignores the fact that compliers in the treatment group are self-selected group of individuals and they often differ from the rest. CACE method offers comparison of observed compliers in treatment group versus potential compliers in the control group. Thus, it estimates effect of treatment based on two comparable groups that were randomly assigned to treatment and control [152].

There are five key assumptions underlying CACE approach [151]: (1) Stable unit treatment value assumption (SUTVA) – potential outcomes for individual *i* do not depend on the treatment status of others, i.e. causal effects are defined for individuals, and everyone receives the same treatment; (2) Outcome exclusion restriction – given the treatment received, treatment assignment does not affect outcome; (3) Monotonicity – there are no "defiers" (those who do the opposite, i.e. $T_i(1)=0$ and $T_i(0)=1$); (4) Non-zero denominator – there are always some compliers; and (5) Randomization – treatment assignment is based on randomization.

The most questionable assumption among these five assumptions is the second assumption. It would not be unlikely that participants who were randomized into treatment, i.e. facilities that were supposed to receive RBF, but who did not comply, meaning they did not receive performance-based payments or at least did not know that they did, would be demoralized by it [153]. The same health workers would not be demoralized or less so, had they been randomized into control, i.e. facilities that did not receive RBF, because they would not feel left out for example. This assumption is one of the five assumptions that are made implicitly or explicitly in most studies using the CACE approach, although in many cases there is a reason to believe that it could be violated [154]. In this study, sensitivity analysis was conducted to test how well the assumption held up and whether it changed the fundamental conclusions regarding the treatment effect of PBP on motivation factors. The sensitivity analysis was performed using the approach described by Jo [153] that uses pretreatment covariates and the

assumption of additive effect of treatment while relaxing the assumption of exclusion restriction. The additivity assumption is necessary to ensure model identification.

The third assumption holds because those assigned to control were not allowed to receive treatment, thus, they could not defy their treatment allocation. The fourth assumption holds as there are health workers in the treatment group who did actually receive the payments (see section on variables construction). The fifth assumption holds because treatment assignment was randomized. A potential limitation of the CACE method in comparison with ITT is how to define compliers [148, 149, 154]. Effects can differ depending on the definition of compliance.

The main equation for the measurement model is summarized below. The subscripts for facility and health worker are not shown in the measurement model. Thus, the equation is for health worker at a facility level:

$$\boldsymbol{y} = \boldsymbol{\Lambda_y \eta} + \boldsymbol{\varepsilon} \quad (1)$$

$E(\boldsymbol{\varepsilon}) = 0, \qquad E(\boldsymbol{\eta\eta'}) = \boldsymbol{\Phi}, \qquad E(\boldsymbol{\varepsilon\varepsilon'}) = \boldsymbol{\Theta_\varepsilon}, \quad$ and $\boldsymbol{\varepsilon}$ uncorrelated with $\boldsymbol{\eta}$

where $\mathbf{y}$ is the 15 x 1 vector of observed components $Y_1, Y_2, ..., Y_{15}$, $\boldsymbol{\Lambda}$ is a 15x4 matrix of indicator-construct loadings, $\boldsymbol{\eta}$ is a 4 x 1 vector of latent motivation factors, and $\boldsymbol{\varepsilon}$ is 15 x 1 vector of error terms of the observed scores in $\mathbf{y}$. It is assumed that measurement errors are uncorrelated with the latent variables or factors [146]. The mean of measurement errors is zero. In addition, it is assumed that factors are correlated, which is in accordance with the theoretical framework of the study. Error terms are assumed to be uncorrelated,

so $\boldsymbol{\Theta_{\varepsilon}}$ matrix is diagonal. This assumption, also referred to as local independence assumption, implies that no two endogenous variables share a common omitted cause. It is the main assumption in EFA, but could be relaxed in CFA. It could be violated if indicators coming from the same source share a common response bias [146]. Errors are assumed to be homoscedastic. They are also assumed to be nonautocorrelated within facility.

The main equation for the structural model using ITT approach is described below:

$$\eta_{ijk} = \gamma x_j + \beta x'_{jk} + \zeta_{ijk} \quad (2)$$

where η_{ijk} is a motivation factor *i* (*i*=1, 2, 3, 4) for an average health worker *k* in the facility *j*, γ is the effect of treatment allocation ($x_j = 1$ treatment), which is a facility level variable, on factor *i*, $\boldsymbol{\beta}$ is a vector of coefficients for covariates at a health worker level (x_{jk}), and ζ_{ijk} is a residual term. A constant (intercept) is absent from the equation because the variables are deviated from their means.

The two main equations for the structural model using CACE approach for factor *i* are summarized below:

$$\eta_{jk} = \alpha_c^{\eta} + \gamma_c^{\eta} x_j + \lambda^{\eta} x_{jk} + \varepsilon_{jk} \quad (3)$$

α_c^{η} is the expected difference in the outcome for a complier as compared to a non-complier in a facility *j* in control arm; γ_c^{η} is complier-average causal effect, i.e. the effect of treatment x_j on compliers (the effect of treatment assignment on non-compliers is zero under the exclusion restriction assumption listed earlier but which will be allowed to vary as part of the sensitivity analysis); λ^{η} is the coefficient for other covariates at

health worker level; and ε_{jk} is the error term. To test the fundamental assumption of exclusion restriction, the additive effect of treatment assignment was assumed, and thus the equation above does not contain a term for interaction effect of treatment assignment with covariates.

Logistic regression model predicting compliance based on facility- and health worker-level variables is shown below:

$$logit(\pi_{cjk}) = \beta_0 + \beta_1 z_j + \beta_2 z_{jk} \qquad (4)$$

where π_{cjk} denotes the probability of being a complier, β_0 is a logit intercept, $\boldsymbol{\beta_1}$ is a vector of logit coefficients for the facility level variables (z_j) predicting compliance, and $\boldsymbol{\beta_2}$ is a vector of logit coefficients for the individual level variables (z_{jk}) predicting compliance.

All data were analyzed using STATA 12 [105] and MPlus v7 [104]. Test of multicollinearity using the variance inflation factor estimation method for complex survey data was implemented for variables selected as potential predictors for ITT and CACE models [155].

ITT analysis was implemented using weighted least squares function for complex survey data with categorical outcomes. Standard error computations used a sandwich estimator in order to account for non-independence of observations due to cluster sampling by specifying TYPE=COMPLEX in the ANALYSIS command of the input file [104].

CACE analysis was implemented using mixture modeling with cross-sectional data in MPlus by specifying TYPE=MIXTURE in the ANALYSIS command of the input file. Mixture modeling is used with categorical latent variables with unknown sub-population membership, but which is inferred from the data. Maximum likelihood estimation approach robust to non-normality and non-independence of data (MLR) was used in estimating the CACE, following the method described by Jo et al that use CACE in the analysis of the data from the Johns Hopkins School Intervention Study[156]. Maximum likelihood method is generally considered to be more efficient than the instrumental variable approach for estimating CACE [150, 156].

4.4 Results

4.4.1 Study population

A total of 805 health workers from 256 facilities (clusters) in 11 RBF pilot provinces were surveyed in this study. Of the 805 health workers, 430 (53.4%) were from 138 facilities allocated to the treatment arm, and 375 (46.6%) were from 118 facilities that were part of the control arm (Table 4.1). Overall, there were no significant differences in facility characteristics (p-value>0.05). In treatment arm there was a higher proportion of CHCs (45.3% compared to 36.2%), while in control arm there was a higher proportion of BHCs (54.6% compared to 43.8%). But these differences were not statistically significant. Distribution by management type was similar between the two arms of the study. There were also no significant differences between the two groups in facility equipment, infrastructure and drug availability indices.

Distribution by type of health worker as well as other health worker characteristics was also similar in two groups, with the exception of supervisor visit within the past 30 days. In both groups males were more than 50% of health workers, reflecting the situation with women's employment in Afghanistan. Vaccinators made the single largest group (27.3% and 29.4% in treatment and control group, respectively), although they were closely followed by nurses and midwives, including community midwives. Mean number of technical training days outside of one's health facility was practically the same in both arms. Also, they both were characterized by large spread as evidenced by standard deviation estimates in the number of training days over the past year. The only significant difference identified in this analysis was variation in supervisor's visit, which was higher in treatment group (90.1%) as compared to control (83.5%). However, there was no adjustment for this variable in the regression analysis as it was not considered to be a confounding variable. There are three conditions that must be met for a variable to be a confounder [157]. It must:

1. Cause the outcome, i.e. association cannot be chance only,
2. Be associated with exposure, and
3. Not be affected by exposure.

While the supervisor's visit was associated with exposure, there was no proven clear causal link between work motivation and frequency of supervisor's visit, because the variable did not show frequency or purpose of the visit but only whether there was a recent visit from a supervisor for *any* reason. More importantly, higher frequency of supervisory visits was partly due to RBF as it did involve more monitoring on behalf of managing organizations and MOPH [158].

4.4.2 Characteristics by compliance status

As expected, there were several important differences between the group of compliers (n=190), i.e. those who reported having received PBP, and the group of non-compliers (n=237), i.e. those who did not report having received PBP (Table 4.2). First of all, there were significant differences among compliers and non-compliers by type of facility. Compliers came mostly from CHCs (55.0%), while non-compliers came mostly from BHCs (53.2%). Thus, it seemed that either staff of CHC facilities was more informed about their bonus payments or these larger facilities were more likely to have received PBP. There was also a very important and clear difference between compliers and non-compliers by type of management. Among compliers, 46.2% worked in facilities managed by NGOs only, while among the non-compliers only 27.9% came from facilities managed by NGOs. At the same time, less than 1% of compliers worked in MOPH only facilities, while 6.3% of non-compliers came from these facilities.

As for health worker characteristics, it seemed that among compliers the majority were females (53.6%), while among the non-compliers there were only 41.1% of female health workers (Table 4.2). There were also differences by type of health worker. As expected, based in part on the type of performance indicators, among compliers 35% of health workers were midwives or community midwives, while among the non-compliers there were only 23.9% of midwives or community midwives. There were also more doctors among compliers than non-compliers (14.7% as compared to 8.7%, respectively). Surprisingly, there was a higher proportion of vaccinators (32.6%) among non-compliers as compared to compliers (22.5%). DPT3 coverage was one of the eleven payment

indicators for BPHS facilities [68]. Although in both groups it was very high, proportion of health workers reporting a recent visit from a supervisor was statistically significantly higher among the compliers (p-value<0.05). These variables, with the exception of supervisor's visit, were tested as predictors of compliance and several of them were selected for the CACE models described further below. Supervisor's visit was affected by exposure [158] as there was more intense supervision of RBF treatment facilities by implementing NGOs and MoPH. Thus, it was not a pretreatment covariate.

4.4.3 Outcome by treatment group

Based on the distribution of individual items of the work motivation scale, it appeared that there were very few observable differences between treatment and control groups (Appendix B, Table 8.1). In both groups, there was a very low frequency of endorsement (<10%) of negative responses ("Strongly disagree" and "Disagree") for the identified regulation factor items. Although there was a considerably less skewed distribution of the items in the introjected motivation factor, there were no significant differences between the two arms. For amotivation factor, a higher proportion of health workers in control facilities as compared to treatment facilities expressed strong agreement. For example, 7.2% strongly agreed with the statement, "I sometimes feel my work is meaningless", in the treatment group as compared to 13.8% in control group. However, these differences were not statistically significant (p-value >0.05).

Unlike the three factors discussed above, there were significant differences in three out of five external regulation factor items. For example, 42.8% of health workers in treatment

arm as compared to 52.6% in the control arm (p-value<0.05) expressed strong agreement with the statement that they worked in a given facility because of opportunities for promotion. A significantly higher proportion of health workers in control arm strongly endorsed the statement that they were working in a given facility because of opportunities to use their skills (57% as compared to 48.1%, p-value<0.05).

4.4.4 Intention-to-treat analysis

A structural equation model using a weighted least squares function was fit to examine the primary relationship of interest, the effect of intervention on motivation factors (Figure 4.2). The model fit statistics were satisfactory and suggested that the model had an acceptable fit to the data, although improvements could be made (Table 4.3). The standardized loading coefficients were all above 0.40 and were statistically significant (p-value<0.05) (Table 4.4). Importantly, they had the exact same magnitude as in the measurement only model, signifying measurement invariance, which is a prerequisite for examining the effect of treatment (see Chapter 3). The structural model coefficients were near zero (Table 4.5). None of them were statistically significant (p-value>0.05).

Based on the results of the exploratory data analysis, it appeared that gender and health worker type were associated with different motivation factors. The models containing these variables were fit next. Fit indices for the model with health worker gender (Model B) only were similar to those of Model A, the model with treatment only. While there were no gender differences in identified motivation and amotivation factors, it appeared

that external and introjected motivation factors were significantly higher among females as compared to males (p-value<0.05).

The model with health worker type as a predictor of motivation factors (Model C) seemed to have similar fit indices as the previous two models, although it had slightly lower RMSEA. Health worker type was found to be a significant predictor of external and introjected regulation factors. Midwives appeared to have significantly higher external (β=0.26, p-value<0.05) and introjected (β=0.40, p-value<0.05) motivation. At the same time, they appeared to be more amotivated than doctors (β=0.29, p-value<0.05). Community midwives had significantly higher external (0.44, p-value<0.05) and introjected (0.39, p-value<0.05) regulations as compared to doctors. Both vaccinators and CHW supervisors also had higher external and introjected motivation when compared to doctors (p-value<0.05). Similar to midwives, CHW supervisors were also significantly more amotivated than doctors (p-value<0.05). Thus, while there were very few differences in identified regulation (autonomous motivation type) among different types of health workers, health workers with lower levels of training and position had higher controlled types of motivation and in some cases, amotivation.

In Model D, facility type was tested as a predictor of motivation factors. None of the coefficients were statistically significant. In Model E, type of management of the facility was tested as a predictor of motivation factors. Similar to Model D, this variable did not appear to be significantly associated with motivation factors.

Model F, combining health worker type and gender as predictors of motivation, appeared to have similar fit indices as the more parsimonious models (Table 4.3). In this model, the coefficient for gender did not appear as a significant predictor for any of the factors, except the external motivation factor (Table 4.5). Females seemed to have higher external motivation (p-value=0.05) in this combined model, holding the type of health worker constant. Unlike in Model B, gender was no longer significant (p-value<0.05) for introjected motivation factor, although the magnitude of the coefficient was similar to the estimate in gender only model.

Health worker type did not appear to be a significant predictor of identified motivation. It was, however, a strong predictor of external motivation (p-value<0.05), holding gender constant. It was also an important predictor for introjected motivation. For amotivation factor, the coefficient for CHW supervisor remained significant (p-value<0.05) and of the same magnitude as in health worker type only model. For midwife, while the magnitude of the coefficient changed only slightly, it was no longer statistically significant (Table 4.5). This is not surprising as regular and community midwives are all females in Afghanistan, and thus, gender and health worker type in this case are strongly correlated. However, the variance inflation factor showed that overall, there was no collinearity between gender and health worker type.

4.4.5 Complier-average causal effect analysis

As described in the variable construction section, observed compliance was defined as a positive response to the question on whether a health worker personally received PBP. According to this definition, 50.8% of all health workers working in facilities allocated to treatment arm was defined as compliers. Thus, almost half of health workers either did not receive treatment or were not aware of it. These were defined as non-compliers. Given that compliance status could only be observed in treatment group, the observed compliance status was missing for the control group.

As the first step, a model with three comparison groups: (i) controls, (ii) non-compliers in treatment arm, and (iii) compliers in treatment arm. It was implemented using multiple group analysis where a reference category was the compliers group in treatment arm whose intercepts were fixed at zero. Health workers in the control group had positive factor means for identified, external and introjected regulations (p-value<0.05) when compared to health workers classified as compliers. Also, we observed smaller but positive factor means (p-value<0.05) for identified and external regulation factors for non-compliers when compared to compliers. It appeared that non-compliers in treatment arm had higher identified and external motivation.

Next, a model without any pretreatment covariates was fit to the data (Table 4.6). The results of the CACE model with treatment as the predictor and motivation factors as the outcome are provided in Table 4.6. Maximum likelihood estimator with robust standard errors (MLR) was used for this part of the analysis. MLR estimates are robust to non-

normality and non-independence of observations [104]. There is no chi-square goodness of fit test in mixture models. Instead, several models must be fit in order to compare their log-likelihoods and information criteria. Entropy (0.89) approaching one provided evidence that there was a clear delineation of classes. Based on the results of this model it appeared that when comparing compliers in treatment groups with compliers in control group, PBP as implemented in Afghanistan had statistically significant negative effect on identified, external and introjected motivation (p-value<0.05).

A series of models with various pretreatment covariates – facility type, management type, health worker gender, and type of health worker – selected based on the results of exploratory data analysis were added one at a time to the initial model with only treatment and compliance. It appeared that facility type and management type were not significantly associated with motivation factors, although they predicted compliance status. Gender was not associated with any of the motivation factors, except external regulation and this became insignificant once health worker type was added to the model, although the direction of the association was similar with males on average being less externally motivated than females. Likelihood-ratio test and the Bayesian information criterion (BIC) adjusted for sample size were used to select the best fitting model [159]. The results of the final CACE model are provided in Table 4.7 - Table 4.8.

Similar to results of the ITT analysis, health worker type was not a significant predictor of identified regulation, although the coefficient for midwives was large (λ=-0.60) it was not significant at 5% level. Community midwives (λ=0.43, p-value<0.05), vaccinators (λ=0.31, p-value<0.05), and CHW supervisors (λ=0.37, p-value<0.05) were estimated to

have a higher mean external motivation as compared to doctors, holding treatment constant. Midwives (λ=0.63, p-value<0.05), community midwives (λ=0.60, p-value<0.05), vaccinators (λ=0.47, p-value<0.05), and CHW supervisors (λ=0.57, p-value<0.05) were estimated to have a higher mean introjected motivation as compared to doctors, holding treatment constant. For amotivation factor, structural coefficients for midwives and CHW supervisors were positive and significant (λ=0.57 and λ=0.46, p-value<0.05, respectively), indicating that these types of health workers were expected to be more amotivated than doctors, holding treatment constant.

Importantly, the magnitude and significance of coefficients for treatment effect changed very little when other covariates were added to the model. As Table 4.7 shows, it appeared that PBP were negatively associated with identified, external and introjected motivation. The estimated complier-average causal effects for these three factors were negative and highly statistically significant (γ=-2.52, p-value<0.01, γ=-1.29, p-value<0.01, γ=-1.30, p-value<0.01).

Based on the exploratory data analysis, it appeared that type of management, facility type, health worker type or position, and gender were potential predictors of compliance status. Using the observed compliance status as an outcome, a series of bivariate and multivariate logistic regression models based on the generalized estimating equations (GEE) with robust standard errors were fit to the data. Management type and facility type appeared to be significant predictors of compliance status. However, it is recommended that the model predicting compliance status contains more or at least the same number of

covariates as the outcome model. Also, it is advisable that at least some of the covariates in two models overlap [160]. Thus, gender and health worker type were also included in the compliance status model (Table 4.8).

As described earlier, the results of this analysis depend fundamentally on the assumption that non-compliers assigned to the treatment group were not affected by this assignment. While it was reasonable to assume that to be affected by the intervention one had to actually participate in it [154], the study tested this assumption. Importantly, testing of this assumption requires covariates that predict outcome and compliance. The coefficient(s) of the covariate(s) predicting compliance status cannot be zero, and the model must contain covariates that are significantly associated with compliance. As Table 4.8 shows, the model predicting compliance status met these two requirements. Management type was statistically significantly associated with compliance status. The odds of compliance were five times higher in facilities that were managed by MoPH with support of an international agency as compared to MOPH only facilities (OR=5.5, p-value<0.01). The odds of compliance were even higher for facilities managed by NGOs only (OR=6.6, p-value<0.01). Furthermore, we must assume the additive effect of treatment assignment. Thus, the model with pretreatment covariates described above was modified to allow the main effect of treatment to be non-zero for non-compliers.

The model with relaxed exclusion restriction assumption showed that there were statistically significant differences in the effects of treatment between compliers and non-compliers for identified and external regulations, suggesting that the exclusion restriction

assumption was perhaps not realistic in this study, assuming the additive effect of treatment assignment holds (Appendix B, Table 8.3). However, more importantly, this alternative specification of the model did not change the fundamental results of the study. While the magnitudes of CACE estimates were slightly smaller in this model as compared to the model with exclusion restriction, the direction of the coefficients did not change (Appendix B, Table 8.3).

Health workers in treatment group who received treatment had statistically significantly lower identified motivation when compared to their colleagues in control group who would have received it had they been allowed to do it (γ=-2.02, p-value<0.05). External motivation was also lower among health workers assigned to treatment group and who actually received the treatment as compared to their colleagues assigned to control group who would have received it had they been assigned to it (γ=-1.09, p-value<0.05). Complier-average causal effect estimate was also negative for introjected motivation (γ=-0.85, p-value<0.05). Although the direction of the effect of treatment on amotivation did not change once the exclusion restriction assumption was relaxed, it became significant unlike in the model with the OER assumption (γ=0.56, p-value<0.05).

It should be noted that based on information criteria, it appeared that the model with no exclusion restriction assumption fit the data slightly better than a model based on this assumption. Likelihood ratio test also showed that it was a statistically significantly better model than the model based on the traditional CACE assumption (LR=167, 4 d.f.). This implies that treatment had an effect on non-compliers as well as compliers. Although the

treatment coefficients on non-compliers must be interpreted with caution [153], it appeared that non-compliers in treatment arm had statistically significantly higher identified and external motivation when compared to non-compliers in control arm. This suggests that those health workers working in treatment arm who reported not receiving PBP had higher identified and external motivation as compared to those who said they received such payments.

4.5 Discussion

Performance-based financing programs have become increasingly popular for addressing gaps in health system performance. As discussed in the introduction to this paper, external rewards, particularly financial incentives, are seen as key mechanisms for improving health worker motivation. However, this assumption has been questioned by theories such as SDT [19, 51]. Also, as several studies, including systematic reviews, have shown there is insufficient evidence supporting this link between external rewards contingent on performance and work motivation [41, 43, 137, 161].

The results of this study suggest that financial incentives in the form of pay-for-performance do not necessarily improve health worker motivation. At the programmatic level, as the results of the intention-to-treat analysis demonstrate, the pay-for-performance project in Afghanistan had no effect on health worker motivation. While other variables, such as gender and health worker type, were shown to be important predictors of different dimensions of motivation, the RBF intervention did not appear to have significant association with any of the motivation factors. These findings support

the findings of the recent evaluation conducted by the Johns Hopkins School of Public Health [75]. The evaluation did not find significant improvements in most of the indicators of health facility performance as measured by the Balanced Scorecard, including provider satisfaction [75].

However, in cases with a high proportion of non-compliers, i.e. those who did not receive the treatment, the ITT analysis results are often attenuated. Moreover, they reflect the programmatic effect of the intervention, i.e. effect of the program on the outcome in realistic context where some participants receive treatment while others do not. Efficacy of the intervention itself is more accurately assessed through methods of causal inference such as CACE. Given that a large proportion of health workers in this study reported that they did not receive pay-for-performance, it seemed reasonable to conduct analysis using the CACE method. The study did not distinguish between those who actually did not receive PBP and those who did receive them but were not aware of them being PBP, because for motivation, a health worker's perception or awareness seemed to matter more than the actual fact of receiving such payments. Also, it was not feasible to distinguish between the two.

Based on the results of the CACE analysis, it seems that PBP can, in fact, have a detrimental effect on identified, external and introjected dimensions of motivation. The negative effect of PBP on identified motivation supports the hypothesis of the study based on SDT. However, findings of no effect on external motivation at best and negative

effect on external motivation at worst were surprising. Yet, there are several design issues of the RBF program in Afghanistan that may help explain these unexpected results.

First of all, it is possible that the program did not meet the heightened expectations of health workers. The project was widely publicized at its start and health workers were likely to have expected higher rewards than what they received in reality. Approximately one year into the implementation of the RBF program, in the last quarter of 2011, there was a revision of payments for most of the indicators [69]. The payments per indicator had to be substantially increased because MOPH and implementing NGO monitoring activities showed that health workers were not satisfied with the rewards.

Related to the first point, it appears that prior to this revision, PBP as a proportion of base salary (not considering other mandatory allowances) ranged from 5.8% for doctors to 11% for assistant nurses (Table 4.9). However, it is worth emphasizing that this estimation is only in proportion to base salary. In addition to base salary, health workers in Afghanistan, particularly female health workers, are entitled to a number of allowances as described in the National Salary Policy [142]. Thus, male health workers serving in rural areas (which are most of the BPHS facilities) are automatically entitled to 50% increase to their base salary, while their female colleagues are entitled to 100% increase. The hardship allowance is increased further for areas that are considered to be rural (100% and 200% for males and females, respectively) and isolated (125% and 250% for males and females, respectively) [142, 162]. There are other allowances as well.

Although after the revision of per unit payments, the average amount of PBP became higher in proportion to base salary (Table 4.9), it was still small in proportion to the overall monthly salary (base salary plus the mandatory allowances). As Table 4.10 shows, for a midwife it ranged from 16.5% to 4.3%, depending on the location. It is also important to note here that only one province, Balkh, in the current study was from the grade 1 group of provinces. The majority came from grades two and three. Moreover, facilities in the study were overwhelmingly from semi-urban, rural and remote areas. Thus, the overall guaranteed salary for health workers from these facilities was high relative to what PBP provided. In Rwanda, the average increase in staff salaries amounted to 38% as the result of PFP [32]. Health workers who were externally motivated and were thus expected to have higher motivation might have been very well disappointed as their extra efforts did not pay off.

Thirdly, it does not appear that there was much transparency among health workers or oversight by MOPH of how payments were distributed within facilities. In Burundi for example, there was a so called "motivation contract" between the health facility management and the employee that described clearly the tasks of a health worker and the "bonus budget" allocated by the health facility for the health worker if she achieved the set targets [163]. Moreover, the payments were made on a monthly basis. Even more importantly, a clear tool was used to share the "bonus budget" among the health workers [163], and thus, there was transparency and clear understanding by health workers of what they were getting and why. Performance-based payments in such context are more

likely to support intrinsic as well as extrinsic types of motivation. This was not the case with the PBP in Afghanistan [164].

Related to the third point, according to the MOPH Health Economics and Financing Directorate, the within-facility distribution was to be determined by implementing NGOs together with facility management. Thus, there were three types of within-facility distribution schemes in Afghanistan: (1) in proportion to existing salary, (2) based on the contribution to performance indicators, and (3) equal amount to all facility staff. Based on MOPH administrative data, it was estimated that 34.8% of health workers in this study sample came from facilities where within-facility distribution was related to performance [71]. At the same time, there were no clear formulas for estimating performance and contribution of individual staff members and health workers definitely did not have "motivation contracts". Moreover, a number of NGOs changed their distribution schemes during the implementation of the project, making it difficult to measure the modification effect of these widely differing distribution schemes.

Eldridge and Palmer in their review of the literature on PBP in developing countries [41] made an important observation regarding the premises of such programs and why they may not result in improved performance or motivation. According to principles of PBP programs, failure to meet performance indicators results in penalty such as lack of bonus payments. This implies that achievement of these indicators is within the control of the provider [41]. To put it differently, it seems that one of the main reasons for low quality of health services is the lack of effort on the part of the health workers. The constraints on

performance are often not within the control of health facilities, and especially their individual staff members. Consequently, PBP may put health workers in a situation where they are expected to achieve targets that are affected by factors beyond their control, and thus, result in lowered motivation, including the external type.

In Rwanda, where the PBP program seemed to be successful not only in improving quality of health services but also to a certain extent successful in increasing health worker motivation [161], an average of 23% of PBP funds were allocated for overall facility-level investments such as infrastructure and medical supplies [32]. A study of the PBP experience in Rwanda by Kalk et al [161] based on 69 semi-structured interview found that one of the reasons for the positive effects of the program was the fact that part of the funds were used for infrastructure and equipment, thus addressing a major constraint to performance. This finding is also supported by the preliminary findings of a qualitative study of an RBF pilot in Nigeria [165]. According to this study, one of the key factors mentioned by health workers regarding the benefits of the PBP program was the overall improvements in facility infrastructure. These improvements in the working environment were seen to be at least as important as the increases in their individual salaries due to PBP [165]. In Afghanistan, all PBP were spent on individual staff salaries and no additional funds were provided for addressing facility-level performance constraints, such as lack of medical supplies or poor infrastructure.

A fifth point that could also partly explain the negative effect of the PBP program on health worker motivation, including external dimension, in Afghanistan is also based on

findings in Rwanda. One of the negative effects of PBP cited by health staff interviewed in Rwanda was increased workload. With a significant shortage of staff in health facilities the PBP program was often perceived as straining an already overstretched system as health workers complained of being overworked and tired [161]. Similar to Rwanda, in Afghanistan, implementing NGOs and health facility managers who felt pressure to achieve targets were likely to put pressure on their staff to increase the number of ANC or PNC visits in relation to baseline (as bonus payments were calculated in relation to each facility's performance) without an increase in staffing levels or other inputs. While a more detailed study is required, facility records of the total number of new patients do suggest that the average workload of each health worker (defined as the number of new patient visits divided by the total number of staff per day) has increased in treatment facilities as compared to control (p-value<0.10).

Last but not least, as was the case in Rwanda, there were delays in payments that undermined the "mental link" between the performance and the rewards [161, 166]. In Afghanistan, facilities were supposed to receive bonus payments on a quarterly basis, which may have weakened the direct link between performance in a particular month and financial rewards received in the form of a salary and other allowances. In addition, however, in Afghanistan there were up to two month delays, and in some cases longer, between the end of the quarter and the time that health workers actually received their payments [166]. All these reasons might explain the surprising findings of this study where on average health workers who were assigned to treatment and complied with it, i.e. actually received PBP, appeared to have lower external motivation as compared to

their colleagues in the control group who would have complied with treatment had they been assigned to it.

Findings of a negative effect of PBP on identified motivation are less surprising and support the study hypothesis. However, as it was noted in the introduction to this study, according to SDT the effect of external rewards does not always have to be negative [19]. The process of the program design and its implementation as well as existing organizational factors appear to be important in determining the effect of PBP on health worker motivation. Based on the studies in Rwanda and Burundi as well as the literature on work motivation, it appears that the process through which the indicators are chosen and the degree to which the rules of payment distributions are understood and accepted by all health workers determine how successful the PBP programs are in improving staff motivation [54, 161, 163] .

For example, in Burundi, as described earlier, there were clear rules for calculating the bonus due to each health worker and these were known and, more importantly, spelled out in individual health worker contracts [163]. As described in Chapter 3, staff participation influences the way external rewards effect autonomous motivation. In Rwanda, one of the main "themes" that came out of interviews regarding the weaknesses of the PBP was the feeling among health workers that indicators were imposed from outside without understanding of local needs and constraints [161]. Thus, they were perceived as mechanisms for controlling health workers. This, according to SDT,

undermines the feelings of autonomy and has a negative impact on identified motivation [19]. It also partly explains the results of this study.

An important factor to also consider when examining these results is the perception of the medical profession in countries such as Tanzania, Rwanda, Burundi or Afghanistan, which may differ from the way it is perceived in countries such as the USA or Canada. In these countries, the respect of community and family members that the health profession brings is extremely important [43, 48, 92, 161]. In Tanzania, it seemed that even financial rewards were closely connected to the perception of social status because a good salary conferred recognition of the importance of their work [48]. The impact of a widely publicized PBP program on community's perception of health workers and the services they render would be an important factor in understanding the effect of such programs on different dimension of work motivation, particularly the introjected regulation. Unfortunately, this requires additional qualitative research that was not feasible in this study.

The study has several limitations. First of all, it is based on a cross-sectional survey and thus, any causal inferences are tenuous and should be done with caution. Moreover, as there were variations on a quarterly basis in the total amount of PBP each facility would have received, a longitudinal study might have reflected effect of financial incentives on external motivation more accurately. A second limitation is related to measurement of the outcome. Although the work motivation scale demonstrated good psychometric properties, the sub-scale on external motivation did not contain an item that focused

specifically on financial rewards or income. It contained items on promotion, job security, and working conditions. The item on income ("I only work here to get paid") turned out to be a weak item that did not load consistently on any of the factors and therefore, was dropped from the final scale as described in Chapter 3. This item was originally borrowed from a study in Kenya [86]. As it appeared to be a weak item in pretesting of the scale in 2011, it was slightly reworded for the final scale. As a recent qualitative study in Burkina Faso, Ghana and Tanzania demonstrated, this item did not seem to perform well and was also excluded from the final work motivation scale [85].

It may be that financial rewards differ from general extrinsic rewards such as promotion or job security, and that was why this item did not load with other external regulation items. However, the Pearson chi-squared statistic corrected for the survey design showed no statistically significant difference between the two treatment arms in external motivation based on this item focusing specifically on pay (p-value=0.54). This provides some evidence that even when external regulation is interpreted as motivation based on explicitly financial incentives, the effect of treatment was not significant.

There are also limitations related to the analysis method, specifically to CACE. As described in the section on data analysis, there are several key assumptions that make the identification of CACE models possible [153]. However, while they are widely used they are assumptions that can be violated. Unfortunately, methods for testing these assumptions are limited and hard to implement in many cases [154]. In this study, the

violation of the exclusion restriction assumption, one of the assumptions that appeared to be most often violated in practice, was tested. However, the model testing this assumption also had to rely on other assumptions, such as the additive effect of treatment. Still, the fact that the coefficients in both models were in the same direction and of similar magnitude provides some confidence in the conclusions of this study.

PBP are growing in popularity as a mechanism to address challenges related to human resources in health in developing countries. It is assumed that these types of financial incentive programs will improve health worker motivation, which in turn will contribute to better performance. This study focused on the first part of this question, i.e. effect of PBP on work motivation. It hypothesized that motivation was a multi-dimensional construct consisting of four different dimensions. In line with most theories in organizational psychology, it was expected that financial rewards will have a positive effect on external motivation. However, comparing health workers assigned to the treatment arm to those assigned to the control arm, the PBP program in Afghanistan did not have significant effect on external motivation. Several design and implementation issues, some of which are discussed above, offer explanation for these surprising findings. This study contributes to research on Results-Based Financing projects examining specifically the association between PBP and health worker motivation in developing countries, a link which is too often simply assumed.

Table 4.1 Selected characteristics by treatment group

	Treatment (n=430)	Control (n=375)	n	p-value
Structural/ fixed characteristics				
Facility level				
By type of facility				
Sub-Centers	10.8	9.2	72	
BHC	43.8	54.6	491	0.35
CHC	45.3	36.2	242	
By managing agency				
MOPH, without support	3.4	2.8	38	
MOPH, with support	59.7	58.9	488	0.92
NGO only	36.9	38.4	279	
Health worker level				
Gender (male)	52.5	57.1	464	0.18
Health worker type				
Doctor	11.9	10.0	78	
Nurse	23.4	25.1	197	
Midwife	12.4	8.7	80	0.32
Community midwife	17.2	16.9	134	
Vaccinator	27.3	29.4	231	
CHW Supervisor	7.8	9.9	85	
Duration of work in this facility				
One year or less	32.3	26.3	237	
2-3 yrs	27.4	31.7	242	
4-5 yrs	12.9	15.7	115	0.28
6-10 yrs	23.0	19.9	172	
More than 10 yrs	4.4	6.3	38	
Modifiable characteristics				
Facility level				
Equipment functionality	84.0 (15.5)	83.9 (10.3)	805	0.34
Pharmaceuticals & vaccines availability	82.5 (14.4)	82.5 (14.4)	805	0.94
Infrastructure index	56.5 (26.0)	53.7 (28.6)	805	0.48
Health worker level				
Supervisor visit within the past 30 days	90.1	83.5	699	0.04
Employee assessment in the past 12 months	79.5	84.2	658	0.20
Salary payment up-to-date	60.8	64.8	499	0.51
Salary increase in the past 12 months	35.0	32.1	279	0.56
Training outside of the facility in the past 12 months (days)	5.5 (5.7)	5.8 (6.0)	794	0.80

Note: The overall sample size is 805 observations. Data are mean (SD) or proportions, unless otherwise indicated. All estimates are based on weighted data. Test statistics are adjusted for multi-stage stratified cluster sampling using Taylor-linearized variance estimation. Strata with single sampling unit centered at overall mean.

Table 4.2 Selected characteristics by observed compliance group

	Compliers (n=190)	Noncompliers (n=237)	n	p-value
Structural/ fixed characteristics				
Facility level				
By type of facility				
Sub-Centers	10.0	11.2	43	
BHC	35.0	53.2	243	0.03
CHC	55.0	35.6	141	
By managing agency				
MOPH, without support	0.6	6.3	22	
MOPH, with support	53.2	65.8	256	0.00
NGO only	46.2	27.9	149	
Health worker level				
Gender (male)	46.4	58.9	244	0.03
Health worker type				
Doctor	14.7	8.7	43	
Nurse	22.0	25.2	104	
Midwife	15.6	8.7	47	0.01
Community midwife	19.4	15.2	73	
Vaccinator	22.5	32.6	116	
CHW Supervisor	5.8	9.7	44	
Duration of work in this facility				
One year or less	34.2	30.2	134	
2-3 yrs	22.4	32.6	121	
4-5 yrs	13.3	12.6	56	0.25
6-10 yrs	26.4	19.9	98	
More than 10 yrs	3.9	4.7	18	
Modifiable characteristics				
Facility level				
Equipment functionality (mean, SD)	86.2 (10.5)	81.8 (19.8)	427	0.24
Pharmaceuticals & vaccines availability (mean, SD)	83.4 (11.2)	81.7 (17.5)	427	0.85
Infrastructure index (mean, SD)	54.3 (26.1)	58.5 (25.2)	427	0.35
Health worker level				
Supervisor visit within the past 30 days	94.7	85.2	385	0.01
Employee assessment in the past 12 months	80.3	78.4	348	0.74
Salary payment up-to-date	59.0	63.3	263	0.55
Salary increase in the past 12 months	35.7	34.7	150	0.88
Training outside of the facility in the past 12 months (days)	5.9 (5.4)	5.2 (5.9)	423	0.18

Note: The overall sample size is 427 observations as these estimates are restricted to treatment group only. Data are mean (SD) or proportions, unless otherwise indicated. All estimates are based on weighted data. Test statistics are adjusted for multi-stage stratified cluster sampling using Taylor-linearized variance estimation. Strata with single sampling unit centered at overall mean.

Table 4.3 Model fit statistics, ITT

Model	Chi-sq	df	CFI	TLI	RMSEA	RMSEA C.I.
Model A (trx)	274	95	0.93	0.91	0.05	[0.04, 0.06]
Model B (gender)	272	95	0.93	0.91	0.05	[0.04, 0.06]
Model C (type of health worker)	298	139	0.93	0.90	0.04	[0.03, 0.04]
Model D (facility type)	282	106	0.93	0.91	0.05	[0.04, 0.05]
Model E (management type)	277	106	0.93	0.90	0.05	[0.04, 0.05]
Model F (gender & type)	319	150	0.92	0.90	0.04	[0.03, 0.04]

Note: The overall sample size for the models is 803 observations due to one observation missing values on one of the key covariates used as predictors, and one observation missing values for all motivation scale items.

Table 4.4 Measurement model estimates, ITT

		Loading	S.E.
Identified regulation			
intr1	I work in this job because I have a chance to help other people through my work	0.66	0.04
intr3	I feel I should personally take the credit or blame for the results of my work on this job	0.58	0.04
intr7	I feel a very high degree of personal responsibility for the work I do on this job	0.68	0.05
intrj4	I work in this job to gain God's grace	0.69	0.06
External regulation			
extrg2	I work here because of opportunities for promotion	0.49	0.04
extrg3	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)	0.67	0.03
extrg4	I work here because it is located in a safe area	0.47	0.03
extrg6	I work here because it provides long term security for me	0.54	0.04
intr5	I work in this job because it allows me to use my skills	0.67	0.03
Introjected regulation			
intrj1	I work here because it makes me feel important	0.73	0.04
intrj2	I do this job because my family would be disappointed if I quit	0.70	0.04
intrj3	I do this job because it gives me respect in the community	0.44	0.04
Amotivation			
amot1	I frequently think of quitting this job	0.63	0.04
amot2	I sometimes feel my work here is meaningless	0.81	0.03
amot3	I don't care much about the quality of work here	0.70	0.04

Table 4.5 Structural model estimates, ITT

Model A (Treatment)			
	Coeff	**S.E.**	**p-value**
Identified regulation	0.01	0.08	0.92
External regulation	0.03	0.07	0.71
Introjected regulation	-0.06	0.07	0.38
Amotivation	-0.07	0.07	0.34
Model F (Type of health worker and gender)			
	Coeff	**S.E.**	**p-value**
Identified regulation			
Nurse	-0.05	0.13	0.72
Midwife	-0.24	0.17	0.16
Community MW	0.16	0.17	0.36
Vaccinators	-0.16	0.13	0.21
CHW Supervisor	-0.09	0.14	0.51
Male	0.09	0.10	0.36
External regulation			
Nurse	0.24	0.10	0.02
Midwife	0.11	0.16	0.49
Community MW	0.28	0.14	0.04
Vaccinators	0.28	0.11	0.01
CHW Supervisor	0.34	0.12	0.01
Male	-0.18	0.10	0.05
Introjected regulation			
Nurse	0.17	0.11	0.12
Midwife	0.28	0.16	0.08
Community MW	0.28	0.14	0.04
Vaccinators	0.27	0.11	0.01
CHW Supervisor	0.35	0.13	0.01
Male	-0.13	0.09	0.16
Amotivation			
Nurse	0.10	0.11	0.37
Midwife	0.22	0.15	0.13
Community MW	-0.08	0.14	0.54
Vaccinators	0.14	0.11	0.20
CHW Supervisor	0.27	0.12	0.02
Male	-0.08	0.08	0.36

Table 4.6 CACE model with treatment and compliance only

AIC			25367	
BIC (sample-size adj)			25465	
Loglikelihood			-12618	
Entropy			0.89	
			Compliers	**Non-compliers**
Final class proportions based on estimated posterior probabilities			0.39	0.61
	Identified	**External**	**Introjected**	**Amotivation**
Treatment	-2.63**	-1.40**	-1.37**	0.39
Intercepts	2.71**	1.54**	1.59**	-0.62

Note: NC – non-compliers, reference class, their intercepts are fixed at 0.
** statistical significance at 5% level.

Table 4.7 CACE model with pretreatment covariates

AIC			25267	
BIC (sample-size adj)			25411	
Loglikelihood			-12539	
Entropy			0.89	
			Compliers	**Non-compliers**
Final class proportions based on estimated posterior probabilities			0.39	0.61
	Identified	**External**	**Introjected**	**Amotivation**
Treatment	-2.52**	-1.29**	-1.30**	0.45
Nurse	-0.09	0.23	0.28	0.16
Midwife	-0.60	0.27	0.63**	0.57**
Community MW	0.09	0.43**	0.60**	-0.11
Vaccinator	-0.32	0.31**	0.47**	0.28
CHW Supervisor	-0.14	0.37**	0.57**	0.46**
Intercepts	2.60**	1.44**	1.51**	-0.68

Note: NC – non-compliers is reference class, their intercepts are fixed at 0.
** statistical significance at 5% level.

Table 4.8 Predictors of compliance status, CACE

	Beta	S.E.	OR	p-value
BHC facility	0.04	0.41	1.04	0.93
CHC facility	0.63	0.41	1.88	0.12
MOPH with support	1.70	0.57	5.47	0.00
NGO only	1.89	0.59	6.59	0.00
Male	-0.10	0.24	0.91	0.69
Nurse	-0.22	0.29	0.80	0.44
Midwife	0.20	0.38	1.22	0.60
Community MW	0.08	0.36	1.08	0.83
Vaccinator	-0.10	0.29	0.91	0.73
CHW Supervisor	-0.18	0.35	0.84	0.61
Intercept C1	-2.24	0.70		0.00

Table 4.9 Base salary levels and average PFP by type of health worker, 2011-2012

	2011			2012		
	Base salary	Average PBP	% of base salary	Base salary	Average PBP	% of base salary
Doctor	8,500	494	5.8	10,700	1,465	13.7
Nurse	5,500	494	9.0	6,900	1,465	21.2
Assistant nurse	4,500	494	11.0	5,700	1,465	25.7
Midwife	7,000	494	7.1	8,900	1,465	16.5
Community MW	6,250	494	7.9	6,300	1,465	23.3
Vaccinator	5,000	494	9.9	5,200	1,465	28.2
CHW Supervisor	6,000	494	8.2	6,300	1,465	23.3

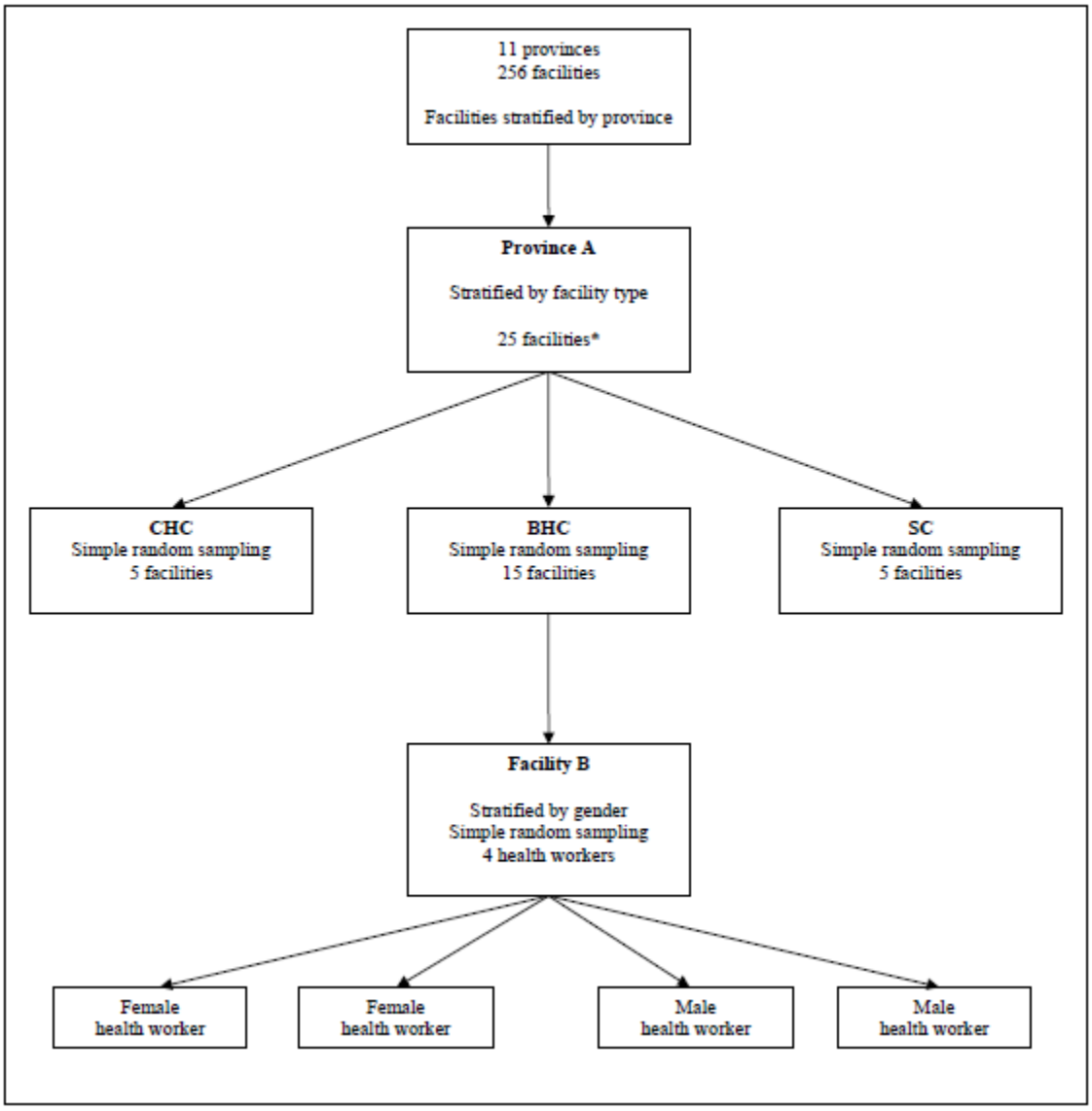
Source: MOPH, 2005; MOPH, 2011; MOPH administrative data.
Note: In NSP 2005 salaries were expressed in US dollars. Conversion rate is 50 Afs=1 USD.

Table 4.10 Base salary, allowances and the average PFP for midwives, 2012

	Urban (0%)	Semi-urban (50%)	Rural (100%)	Deep rural & remote (200%)	Isolated (250%)
Base salary	**8,900**	13,350	17,800	26,700	31,150
Grade 1 province	8,900	13,350	17,800	26,700	31,150
Grade 2 province	9,782	14,232	18,682	27,582	32,032
Grade 3 province	10,663	15,113	19,563	28,463	32,913
Grade 4 province	11,545	15,995	20,445	29,345	33,795
PFP as % of salary					
Grade 1 province	16.5	11.0	8.2	5.5	4.7
Grade 2 province	15.0	10.3	7.8	5.3	4.6
Grade 3 province	13.7	9.7	7.5	5.1	4.5
Grade 4 province	12.7	9.2	7.2	5.0	4.3

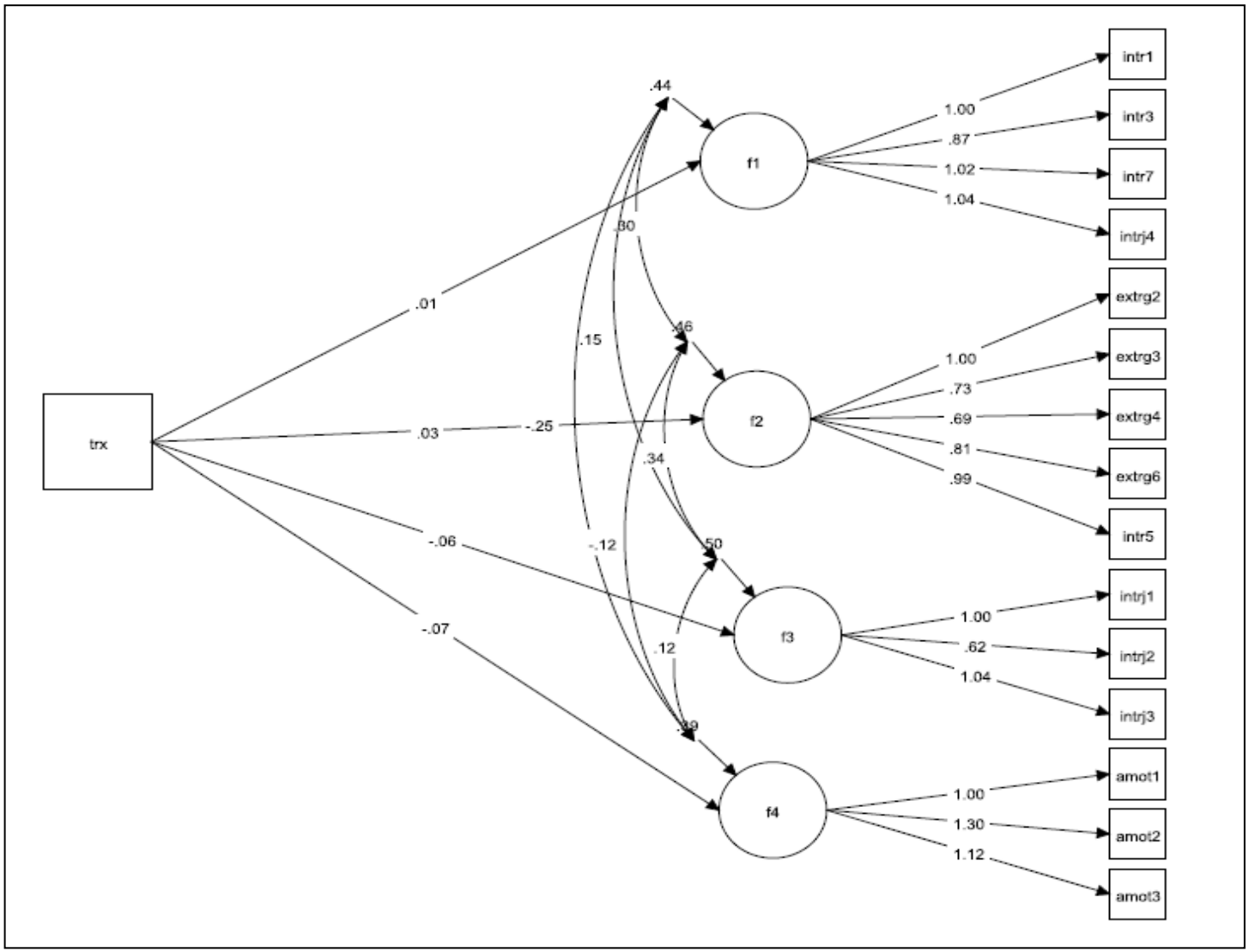
Source: NSP 2011; MOPH administrative data.

Figure 4.1 NHSPA 2012-2013 sampling for health worker questionnaires



Note: * In provinces with less than 25 facilities, all facilities have been surveyed. In Kandahar, where only eight facilities were selected for RBF pilot, only eight facilities were included in RBF sample. In the following RBF pilot provinces – Parwan, Takhar, Kunduz & Saripul – more than 25 facilities per province were selected to ensure inclusion of matched pair facilities surveyed at baseline in 2010-2011.

Figure 4.2 Pay-for-performance and health worker motivation (f1 – identified, f2 – external, f3 – introjected, f4 – amotivation)



5 Chapter 5: Impact of performance-based payments on quality of care in Afghanistan (Paper 3)

Abstract

Background: Progress towards achieving Millennium Development Goals on maternal and child health has been slow. Performance-based payments for health care providers showed promise as a strategy for increasing utilization and quality of health services for women and children. This study examines the effect of performance-based payments on quality of care in Afghanistan.

Method: In 11 provinces of Afghanistan, 442 facilities were randomly assigned to an intervention or a comparison arm. Because of the study design, masking of the participants was not feasible. The study is based on a cross-sectional survey of 2,180 patients and 255 health workers from 233 facilities conducted 23 months after the start of the intervention. The outcome, quality of care, was measured using a standardized checklist. The list consisted of 16 items designed to measure four areas of clinical quality of care: patient history, physical examination, patient counseling, and time spent with patient. The main approach to the analysis of the data was a linear mixed effects model using maximum likelihood estimator. In addition to model-based standard errors, robust standard errors using the sandwich estimator were obtained as part of the process of residual diagnostics. The effect of treatment was estimated using intention-to-treat analysis.

Results: Based on the likelihood ratio test, it was determined that there was insufficient evidence to justify a three-level model, thus a two-level model with a random intercept for each health worker was fit to the data. The model with no covariates showed that 55% of total variation in quality of care was attributable to differences across health workers, while 45% of the variation in the outcome was due to within health worker differences. Comparing patients in the treatment group to patients in the control group, the expected change in z-score for quality of care was 0.23 (p-value<0.05). The decrease in level-2 variance provided further evidence that treatment explained some of the variation in the outcome observed between health workers. Other covariates were added to the model. Together, they explained 15% of the total variance in quality of care, while the proportion of level-2 variance explained by the covariates was 26%.

Conclusion: The study demonstrated that the PBP program had a significant positive effect on quality of care. Moreover, it explained some of the observed heterogeneity in quality of care between health workers. At the same time, the study showed that good management practices such as paying salaries on time and ensuring presence of proper equipment are important predictors of quality of care. Moreover, they explain a higher proportion of total variance in observed quality of care than PBP. Also, the study found a significant positive association between identified motivation and quality of care, holding treatment constant. However, further research linking specific organizational factors that promote identified motivation such as ability of staff to voice their opinions and participate in decision-making to quality of care is needed. Also, lack of association

between external motivation and quality of care found in this study needs to be explored further, particularly as PBP projects target this dimension of motivation.

5.1 Introduction

Despite significant amounts of development assistance for health and education provided to Afghanistan over the past decade, it still has some of the world's highest maternal mortality ratios and child mortality rates. According to The State of the World's Children 2012 report [7], among 193 countries ranked from highest to lowest, Afghanistan is ranked number eleven in under-five mortality rate. It was estimated that in 2010 it had 149 deaths among children less than 5 years of age per 1000 live births compared to Iceland that had 2 [7]. According to the systematic analysis of progress towards Millennium Development Goal 5 (MDG 5) by Hogan et al [3], in 2008, Afghanistan had the highest maternal mortality ratio (MMR) of 1,575 maternal deaths per 100 000 live births (uncertainty interval 594–3396), which was about 394 times higher than Italy, the country with the lowest MMR of 4 (3–5). It was also one of the countries with the slowest yearly rate of decline in MMR [3]. Thus, the government and development partners were in search of new interventions to improve access and quality of maternal and child health services in Afghanistan.

In the summer of 2010, Afghanistan launched an RBF project in purposefully selected 11 of its 34 provinces (for details on RBF design see Chapter 1 and [68]). The project aimed "to impact MDGs 4 and 5 by improving coverage of maternal and child health services within the existing health system and without creating unnecessary parallel processes" [10]. Its primary objective was "to increase key maternal and child health outputs, to improve quality of health services and to ensure higher patient satisfaction with health service delivery" [10]. It was hypothesized that improved financial incentives in the form

of pay-for-performance would improve health worker motivation and satisfaction, which in turn would lead to better performance.

Performance-based payments for health care providers showed promise as growing evidence from countries such as Rwanda and Haiti seemed to suggest its effectiveness in increasing utilization and quality of health services for women and children [8, 9]. Moreover, this approach seemed to be supported by theory. According to the World Bank Development Report, Making Services Work for Poor People, weak incentives for performance were viewed as one of the key reasons why health services failed the poor. PBP were thought to be one of the mechanisms of increasing provider accountability to their patients or rather clients, and thus, improving quality of health care services for them [2]. According to Fritsche and Vergeer [163], "Performance-based financing is a health systems approach with an orientation on results defined as quantity and quality of service outputs. This approach entails making health facilities autonomous agencies that work for the benefit of health related goals and their staff." This view of PBP as a mechanism for improving responsiveness of health providers to the needs of their clients, creating greater accountability for results, and empowering consumers is also described in Meessen et al [8].

However, a review of the literature by Eldridge and Palmer [41] did not find strong evidence base for the success of PBP programs in developing countries. This view is supported by a more recent review by the Cochrane Collaboration of PBP in low- and middle-income countries. According to it, there is limited robust evidence of the impact

of PBP and findings vary in direction as well as magnitude of change [62]. The aim of this paper therefore, is to assess quality of care in health facilities allocated to performance-based payments as compared to health facilities allocated to control in Afghanistan.

5.1.1 Research questions

The study aims to answer the following research questions:

3. What is the effect of PBP on clinical quality of care?
    a. Is there substantial heterogeneity in quality of care between facilities and health workers that is unexplained by treatment?
    b. Do facility and health worker characteristics other than PBP explain part of the observed variation in quality of care across facilities and health workers?
    c. Is there evidence of effect modification of the effect of treatment by facility and health worker characteristics?
4. Do motivation types have an effect on quality of care and explain part of the observed variation between health workers, after accounting for treatment effect?
5. Do patient level characteristics explain some of the observed variation in quality of care, after accounting for treatment and other facility and health worker characteristics?

5.2 Methods

5.2.1 Data source

The data for this study was derived from the cross-sectional health facility survey for the annual National Health Services Performance Assessment (NHSPA) implemented in Afghanistan over the period of nine months between 2012 and 2013. The survey was based on a multi-stage stratified probability sampling approach where primary sampling unit was a health facility (Figure 5.1):

1. Stratification by province,
2. Stratification by type of health facility within each province,
3. Selection of facilities,
4. Selection of individual health workers and selection of patients.

Individual health workers and patients were secondary sampling units. Within each facility four health workers classified as clinical staff (doctors, nurses, assistant doctors, midwives, community midwives, and vaccinators) present at the day of the survey were randomly selected for the survey. If fewer than four clinical staff were present at the facility, then other facility staff members, including community health supervisors, pharmacists, cleaners and others, were interviewed for the survey. However, all non-clinical workers with exception of community health supervisors were excluded from this analysis.

Patients were selected independently of health workers. Within each sampled facility, patients were stratified by age group (less than five years of age and five years of age or older) and selected based on a systematic random sampling approach as described below:

- If less than 10 new patients were expected in a day in each age stratum, then each eligible patient was selected until 5 observations of consultations involving patients in that age stratum were completed;
- If 11 to 15 new patients were expected in a day in each age stratum, then every second eligible patient was selected;
- If 16 to 20 new patients were expected in a day in each age stratum, then every third eligible patient was selected;
- If more than 20 new patients were expected in a day in each age stratum, then every fourth eligible patient was selected.

This resulted in a sample of 233 health facilities with 255 health workers, and 2,180 patients. On average, there are 1.1 health workers per facility, 8.5 patients per health worker, and 9.4 patients per facility.

5.2.2 Variables construction

Outcome variables

Quality of care

A composite index of quality of care was developed using 16 items designed to measure four areas of clinical quality of care: (i) Patient history, (ii) Physical examination, (iii) Patient counseling or communication, and (iv) Time spent with patient. The items reflect the MOPH standards of care as outlined in the Guidelines for the Basic Package of Health Services (BPHS) [66]. They have been used as part of the NHSPA since 2004 [40]. Moreover, this index was used in the study by Hansen et al assessing quality of care in Afghanistan [35], although several items have been revised in 2011 at the request of the MOPH as part of the general revision of the NHSPA instruments (see [40] for more details).

The items were measured through observations of patient-provider interactions by a trained observer with a medical degree (forms F1 and F2). Observers were given a standardized checklist with responses coded as "Yes" if a health worker performed a particular action and "No" otherwise. Time spent with patient was estimated as the difference between the time (in hours and minutes) the consultation started and the time it ended. The observer was also asked to record separately the duration of the consultation and this was used as a check for the estimated time.

The items were grouped into three sub-components, measuring (a) patient history and physical assessment consisting of seven items, (b) patient counseling consisting of eight

items, and (c) time spent with patient consisting of one item. Time spent with patient was made into a binary variable. It was coded as 1 if the provider spent 10 minutes or more with the patient and 0 otherwise. This cut-off was estimated as the minimum time needed to complete a history, physical examination, diagnosis and provide counseling for a child with a common respiratory or gastrointestinal complaints [167]. These sub-components are used to estimate separate indicators in the Balanced Scorecards (BSC) for performance of health facilities in Afghanistan [40]. Following Hansen [35], the composite index of quality of care was constructed by averaging the scores on these three sub-components with each sub-component given an equal weight. The composite index was standardized at grand mean to have a mean of zero and a standard deviation of one (z-score).

Predictor variables

Treatment

This is the main predictor variable that is measured at a cluster (facility) level. It is a binary variable based on a random allocation of facilities to either treatment or control at the start of the RBF program. Facilities were assigned to either treatment (coded as 1) or control (coded as 0) based on the administrative data provided by MOPH.

Facility type

Different types of facilities differ in staffing norms, infrastructure, services that they are mandated to perform and size of the catchment population [32]. Thus, this is a good proxy variable for the size of the facility and level of infrastructure. SC is the most basic

type of health facility included in RBF. Each SC serves about 3,000 – 7,000 people. It should have 1 male nurse and 1 female community midwife [32]. BHC has a catchment area of 15,000 – 30,000 people and has the following staffing norms: 1 community health supervisor, 2 vaccinators, 1 male nurse, 1 community midwife, 1 general physician, preferably female [32]. CHC has a catchment area of 30,000 – 60,000 people and has the following staffing norms: 1 community health supervisor, 2 vaccinators, 1 male nurse, 1 female nurse, 2 community midwives, 1 male general physician, 1 female general physician, 1 laboratory technician, 1 pharmacy technician, and one (male or female) psychosocial counselor [32]. CHC provides specialized outpatient care and basic inpatient care. Based on the original categorical variable, dummy variables were created for the analysis where SC was used as a reference group. This variable came from the facility form F7.

Management type

There are three main types of management of health facilities in Afghanistan: (1) MOPH only, (2) MOPH with institutional support, and (3) NGOs only. The latter is further sub-divided by type of contracting which is not one of the study variables here. Facilities run by MOPH only, MOPH with support, and NGOs only differ in management structure, source of funding, and its levels [33]. While there are varying degrees of managerial autonomy within the third group, facilities in this group have the highest level of independence when compared to the other two groups [34]. Thus, it is an important variable containing information on many aspects of facility management – such as funding level, organizational structure and culture, ability to take timely decisions on

things like staffing, system of rewards, infrastructure – that were not otherwise measured through this survey. This variable came from the facility form F7.

Other health facility characteristics

The equipment functionality index consists of 20 items for BHCs and SCs and 23 items for CHCs. It includes items such as children's scales, sterilizer, stethoscope and others that are mandated by the BPHS Guidelines[66]. The composition of the index was based on the methodology provided in the BSC used in Afghanistan for evaluation of health system performance since 2004 [40]. An item is rated as 1 if it is available and is in working condition, and 0 otherwise. Following this, the scores for each item were added together and divided by the total number of items (20 for BHCs and SCs and 23 for CHCs). The variable was standardized as a z-score at a grand mean. Due to its highly negatively skewed distribution it was transformed (original^3) prior to testing of differences in means between the two treatment groups [103]. The index was measured at the facility level.

The vaccines and pharmaceuticals availability index consists of 31 items that are considered to be the essential medicines and vaccine that must be available at all BPHS facilities. Similar to the equipment functionality index it was based on the methodology provided in the BSC revised by the national technical working group led by MOPH [40]. The item is rated as 1 if it has been continuously available for the past 30 days, and 0 otherwise. The variable was standardized as a z-score at grand mean. Due to its highly negatively skewed distribution it was transformed (original^3) prior to testing of

differences in means between the two treatment groups [103]. The index was measured at the facility level.

This index consists of nine items that describe presence and condition of basic infrastructure in the facility, such as heat, electricity, water source, windows etc. The original index used in the BSC [40] was composed of ten items. However, following the same methodology as in the previous Chapter presence and condition of exterior walls was removed from the index. This particular item seemed to indicate less about the quality of services at facilities than items such as reliable electricity and water supply, or presence of interior walls and windows, and it had high proportion of missing values [4]. The item is rated as 1 if is present and few or no repairs are needed, and 0 otherwise. The variable was standardized as a z-score at a grand mean. It was approximately normally distributed.

Health worker type and gender

A self-administered health worker questionnaire contained questions on basic health worker characteristics (form F5), including gender and type of health worker. Gender is a binary variable, where 1 means male. Health worker type is a categorical variable. Health workers were asked to mark one of the following categories: doctor, nurse, midwife, community midwife, CHW supervisor, vaccinator, or other. The last category included administrator, laboratory technicians, pharmacists, guards, cleaners, and other non-clinical staff. These were excluded from the current study. The categorical variable was transformed into dummy variables where doctors were a reference group.

Health worker motivation types

Health worker motivation types are measured through four factor scores reflecting multi-dimensional nature of motivation. As a latent variable, motivation was measured through a self-administered 15-item scale (form F5). As described in Chapter 3 in detail, using exploratory factor analysis followed by the confirmatory factor analysis, the factorial structure of the scale was identified. The four factors of the scale are: (1) identified motivation consisting of four items, (2) external regulation consisting of five items, (3) introjected regulation consisting of three items, and (4) amotivation consisting of three items (see Chapter 3 for definitions of these factors and details of the scale). Given that responses were measured on ordinal scale, the model was fit to a polychoric correlation matrix using the method of robust weighted least squares [104]. As recommended by Skrondal [168] and using the procedures described in MPlus Guide [104, 169] for categorical variables with the weighted least squares estimator, factor scores were estimated as the maximum of the posterior distribution of the factor, or the Maximum A Posteriori (MAP) method (see [168, 170, 171] for discussion of different methods for estimating factor scores and their use in further analysis in regression models). The estimated factor scores are measured on a continuous scale. To help with interpretation of the slope and given that the primary relationship of interest in this paper is between level-2 variable (treatment) and level-1 (outcome) [172], the factor scores, similar to other continuous variables used in the analysis were standardized at grand mean.

Other health worker level variables

Three other health worker level variables were tested as predictors of the quality of care. These are (i) salary payment up-to-date, (ii) knowledge score, (iii) number of training days outside of the facility. These variables are considered to be modifiable in that they could have been influenced by PBP program (treatment variable) and hence, theoretically should not act as confounding variables. However, the design of this particular study did not include increased training for health workers or particular actions targeting regular salary payment processes.

Health workers were asked whether their salary payments were up-to-date (form F5). The answers were coded as "Yes" and "No". Based on the response to this question a binary variable was created where one means "Yes", i.e. there were no delays in salary payments. Health worker knowledge score was based on a knowledge test questions designed as part of the NHSPA for the BSC [40]. The number of questions and their level of difficulty depended on health worker type and ranged from thirty six for doctors and nurses to twelve for CHW supervisors. As part of the health worker questionnaire, health workers were also asked about technical training they received over the past 12 months outside of the health facility. The response was a continuous variable reflecting the total number of training days. It had high positive skew and was transformed through the square root that is often used with positively skewed variables [103].

Patient age and gender

As the performance indicators under the RBF scheme targeted services for women and children under five years of age, it seemed reasonable to assume that quality of care may depend on patient's age and gender. Both variables were recorded during the patient-provider observations. Gender is a binary variable, where one means male. Patient's age was a continuous variable that was transformed into a binary variable where one means five years of age or older. It was recorded only for those patients whose age was asked by the provider. Thus, there were 85 observations with missing age variable. However, since two different forms (F1 and F3) were used for patients less than five and those five years of age and older, it was assumed that patients whose observations were recorded in F1 were less than five years of age and those whose observations were recorded in F3 were five years of age or older. In addition, age was checked against the information given by patients and their caretakers in exit interviews recorded in forms F2 and F4.

5.3 Analysis

5.3.1 Missing data

With the exception of the knowledge score questions, there were very few missing values for the variables used in the study (<5%). None of the 16 items in the quality of care index had more than 1.7% of missing values. Proportion of missing values in any of the items of the index was 3.4%. Given the small proportion of missingness no imputations were performed for these variables.

For health worker knowledge score, out of 75 doctors, nine had a missing value on at least one of the 36 questions. Out of 153 nurses, 29 had a missing value on at least one of the 36 questions. Out of twelve midwives, two had a missing value on at least one of the twenty seven questions. Out thirteen community midwives, two had a missing value on at least one of the twenty seven questions. While overall there were few missing values for each item, because the score consisted of the average of a large set of items (ranging from thirty six for doctors and nurses to twelve for CHW supervisors), a missing value on even one question would result in listwise deletion of all the other responses. This would lead to a high proportion of missing values for the score as a whole. Thus, if a health worker did not respond to a particular question, it was assumed that she did not know the answer and it was coded as zero. As the answer choices did not have "Don't know" option, it seemed reasonable to assume that health workers who skipped particular questions could not respond to them.

However, as part of the sensitivity analysis, two additional methods were used for estimating models with the knowledge test score: (i) multiple imputation technique, accounting for clustering of the data and using procedures described in Hamilton [103], and (ii) complete case analysis, i.e. using only those observations that had no missing values for any of the items of the knowledge test. Complete case analysis was likely to provide biased results as it appeared that missingness was related to health worker type (Table 5.1). Also, it seemed to be unequally distributed between males and females (13.6% and 27.4%, respectively). Thus it was not missing completely at random (MCAR), the only case for which complete case analysis provides unbiased results.

5.3.2 Analysis approach

The main approach to the analysis of the data was a linear mixed effects model using maximum likelihood estimator. Restricted ML (REML) estimator was used to get unbiased variance estimates, which were compared to MLE results. As expected with large samples, MLE and REML provided very similar estimates. In addition to model-based standard errors robust standard errors using the sandwich estimator were obtained as part of the process of residual diagnostics. As described earlier, the data had three levels: facility (level-3), health worker (level-2), and patient (level-1). However, there were only 1.1 health workers per facility. Thus, it was likely that clustering at the facility level was insignificant, which would reduce the three-level data to a two-level data. This was tested using likelihood ratio test, as described in detail further in the section. The equations for the models summarized below assume a two-level data where *j* denotes level-2 observations (health worker or facility) and *i* denotes level-1 observations (patients):

$$QoC_{ij} = \beta_0 + \zeta_j + \varepsilon_{ij} \quad (1)$$

Where QoC_{ij} is the expected quality of care z-score for patient *i* in a facility *j*, ζ_j is the random deviation of a facility j's mean measurement from the overall mean β, which is also an average z-score for quality of care for a typical facility, and ε_{ij} is level-1 residual.

$$QoC_{ij} = \beta_0 + \zeta_j + \beta_1 Trx_j + \varepsilon_{ij} \quad (2)$$

Where β_0 is the expected z-score for quality of care for patients from a typical facility (i.e. facility with ζ_j =0) from the control arm, ζ_j is facility specific deviation from β_0, β_1

is the expected change in z-score for quality of care for patients in treatment arm as compared to patients in the control arm, and ε_{ij} is the level-1 residual.

$$QoC_{ij} = \beta_0 + \zeta_j + \beta_1 Trx_j + \sum \beta_2 \boldsymbol{x_j} + \sum \beta_3 \boldsymbol{x_{ij}} + \varepsilon_{ij} \quad (3)$$

Where β_0 is the expected z-score for quality of care for female patients from a typical facility (i.e. facility with ζ_j =0) from the control arm, β_1 is the expected change in z-score for quality of care for patients in treatment arm as compared to patients in the control arm, controlling for other covariates, $\boldsymbol{x_j}$ is a vector of level-2 potential confounding variables, and $\boldsymbol{x_{ij}}$ is a vector of level-1 potential confounding variables, including gender (where male=1). Continuous level-1 covariates are centered at the grand mean as the main question of interest is the effect of a level-2 dummy variable (treatment), controlling for individual differences on level-1 covariates [172].

Key model assumptions:

1. $\mathrm{E}(\varepsilon_{ij}|\boldsymbol{X_j}, \zeta_j) = 0$

   This assumption implies that level-1 residuals have a mean of zero, given the covariates and the random intercept (level-1 exogeneity).

2. $\mathrm{E}(\zeta_j|X_j) = 0$

   This assumption implies that there is no correlation between covariates and the random intercept (level-2 exogeneity).

3. $Var(\varepsilon_{ij}|X_j, \zeta_j) = \theta \quad Var(\zeta_j|X_j) = \psi$

   It is assumed that variances of level-1 residuals and random intercepts are homoskedastic and are normally distributed.

In general in randomized controlled trials there is no problem of endogeneity due to omitted covariates [159]. Assumption 2 where endogeneity at cluster-level could be due to covariates that have different within-effects as compared to the between-effects was examined using Hausman's endogeneity test as described further below. Assumption 3 was examined using residual diagnostics tools the details of which are described further in this section.

First, all study variables were checked for consistency. Dataset containing information on patient-provider observations were merged on facility ID and unique health worker tracking number. Health worker gender and type (e.g. doctor) were compared between two datasets to ensure that health workers that merged based on their tracking numbers were indeed the same health workers. Where there were differences in either gender or health worker type variable between the two datasets, even though they merged based on health worker tracking number, further investigation was conducted by for example, comparing this information to facility records (form F7) that provides a list of its registered staff by type and gender. Health worker gender values were corrected based on health worker type – all midwives and community midwives were coded as females because according to BPHS Guidelines these positions must be filled by female staff [66]. After these corrections, only those health workers that merged, had the same gender and were coded as the same type were kept in the final dataset.

All categorical variables were examined using frequency distribution tables and bar graphs. Simple summary statistics (means, medians, standard deviations and maximum

and minimum values) were used to examine continuous variables. Stem-and-leaf plots, histograms and box plots were used to do visual checks and examine the symmetry, center, spread and outliers for continuous variables. Outliers were first checked using box plots and more formally with "lv" function in STATA [103]. Values above or below the inner fences are considered to be mild outliers and values above or below the outer fences are considered to be severe outliers [103].

Histograms, box plots and scatter plots were used to examine distribution of quality of care index by the following variables: (a) treatment group, (b) type of facility, (c) management type, (d) three facility-level indices of structural quality (equipment, pharmaceuticals and vaccine availability, and infrastructure), (e) health worker gender, (f) health worker type, (g) number of training days, (h) knowledge score, (i) up-to-date salary payment, and (j) four factor scores of motivation. A graph of cluster means and their corresponding 95% CI was used to examine the heterogeneity in quality of care index (QoC) across clusters as compared to within clusters.

Following the exploratory data analysis, selected characteristics of the sample were examined by treatment allocation. Means and standard deviations were estimated for continuous variables, while for binary and categorical variables estimates of proportions were used. All estimates were weighted to account for unequal probability of selection as it is recommended for descriptive analysis of population characteristics [143]. Weights were based on probability of facility selection. Test statistics were adjusted for multi-

stage stratified cluster sampling using Taylor-linearized variance estimation using the survey commands in STATA [105].

These variables were divided into two parts: structural or fixed and modifiable characteristics. The so called fixed characteristics were unlikely to be affected by PBP, while variables classified as modifiable could have changed as the result of treatment. If that was the case, controlling for them could incorrectly diminish the effect of treatment. For example, if facilities allocated to treatment invested more into training their health workers or used part of the funds for improving availability of drugs, including them in the model with treatment would be incorrect. However, it appeared that there were no systematic efforts on the part of facilities in PBP group to increase training of their staff, and the design of the project did not envision the use of PBP for facility-level expenditures.

Motivation factors, unlike other modifiable characteristics, were targeted by the intervention: It was hypothesized that PBP would increase health worker motivation. However, as described in the previous chapter, based on the results of the intention-to-treat analysis it appeared that treatment had a very small and statistically not significant effect on motivation factors. Also, as Table 5.2 shows, the estimated mean factor scores between the two treatment groups were practically identical. Thus, these variables were explored to see if they could explain the observed heterogeneity in the outcome that remained unexplained by the main predictor of interest, PBP.

A three-level intercept only model was fit with health facility as level-3, health worker as level-2, and outcome as level-1variables. However, given the small number of health workers per facility (range 1 to 2, average 1.1) it seemed that a two-level model could be more appropriate with clustering at either a facility or a health worker level only. Thus, a two-level intercept only model was fit with facility as level-2 and outcome as level-1 variables. Next, a two-level intercept only model was fit with health worker as level-2 and outcome as level-1 variables. Each two-level model was compared to the three-level model using the likelihood ratio test [159]. In addition, based on the results of a three-level model the variance at level-3 was compared with variance at level-2.

Following this, treatment variable was added to the null model. Magnitude of the coefficient and its statistical significance ($p<0.05$) were examined. Also, coefficient of determination (R^2) was estimated to determine if the treatment variable contributed to the proportional reduction in prediction error variance [159]. Other variables were added separately to the model to determine whether the effect of treatment on quality of care was confounded by other variables. A final full model that included all statistically significant variables from previous smaller models was fit to the data. Likelihood ratio tests were conducted to compare it with smaller models. Coefficient of determination was estimated comparing the full model to the model containing only the treatment variable.

Residual diagnostics of the fitted models was conducted to check on the normality assumption for the random intercept and level-1 residuals. According to Rabe-Hesketh and Skrondal [159], while small departures from normal distribution do not pose

problems, severe skewness and outliers signal model misspecification. In cases where histograms of predicted standardized level-1 and level-2 residuals did not appear to be approximately normal, additional check of the model results was done using the sandwich estimator to obtain robust standard errors [159]. Finally, Durbin-Wu-Hausman's test or Hausman's endogeneity test was used to compare random-intercept model with the fixed effects model to see if there was strong evidence for model misspecification, such as correlation of the random intercept with any of the covariates (level-2 endogeneity) [159, 173].

5.4 Results

5.4.1 Study population

A total of 2,180 patients that were seen by 255 health workers from 233 facilities (clusters) in 11 RBF pilot provinces were observed in this study. Of the 2,180 patients, 1,143 (52.4%) were from 123 facilities allocated to the treatment arm, and 1,037 patients (47.6%) were from 110 facilities that were part of the control arm (Table 5.2). Overall, there were no significant differences in facility characteristics (p-value>0.05). In treatment arm there was a higher proportion of CHCs (24.0% compared to 20.8%), while in control arm there was a higher proportion of BHCs (58.2% compared to 49.1%). But these differences were not statistically significant. Distribution by management type was similar between the two arms of the study. There were also no significant differences between the two groups in facility equipment, infrastructure and drug availability indices (Table 5.2).

Distribution by type of health worker as well as other health worker characteristics was also similar in two groups. In both groups, most of the patients were seen by male health workers (87.4% and 90.1% in treatment and control group respectively), reflecting the fact that most patients in the sample were seen either by a doctor or a nurse who are typically male in Afghanistan. As Table 5.2 shows in both treatment and control group, more than 90% of observations came from either a doctor or a nurse. There were no differences between treatment and control group in mean factor scores for the four factors representing four types of motivation. Mean number of technical training days outside of one's health facility and the knowledge score were practically the same in both arms. Also, there were no significant differences in proportion of health workers whose salaries were paid up-to-date and who had a salary increase in the past 12 months. Proportion of patients less than 5 years of age was similar between the two groups, which was to be expected given the sampling scheme. However, there was a significant difference in gender distribution among patients (p-value<0.10). There was a slightly higher proportion of male patients seen in the control group (47.5%) as compared to the treatment group (43.1%).

5.4.2 Exploratory data analysis of the outcome

The data was unbalanced: the number of patients per facility ranged from 1 to 11 with an average of 9 patients. Given that the number of health workers per facility was 1.1 on average, number of patients within each health worker also ranged from 1 to 11 patients with an average of 8.5. As Figure 5.2 shows based on cluster-level average scores for quality of care, it appeared that facilities in treatment arm had higher quality of care

(QoC) as compared to facilities in the control arm. Figure 9.1 (Appendix C) suggests that there is substantial between- as well as within-facility heterogeneity of treatment effect on QoC score. As histograms in Figure 5.3 show, it appeared that outcome was approximately normally distributed. The box plots appeared to suggest that there could be some outliers, particularly in the control group (Appendix C, Figure 9.2). The formal check using "lv" function in STATA showed there were 43 that were considered to be severe outliers (2.0%). They came from 19 facilities and were approximately equally distributed between treatment and control arm.

Next, outcome was examined by different facility, patient and health worker characteristics. As Figure 9.3 (Appendix C) shows, based on cluster-level average scores for QoC, it appeared that facility type affects quality of care: SCs appeared to have the lowest quality of care while CHCs appeared to have the highest, although the difference between BHCs and CHCs did not appear to be large. It also seemed that type of management may affect QoC (Appendix C, Figure 9.4): Facilities managed by MOPH only seemed to have higher quality of care among the three groups. There appeared to be positive association between the equipment functionality index and QoC (Appendix C, Figure 9.5). However, these are based on cluster-level means where the number of observations per cluster is not accounted for.

It appeared that being a male health worker was positively associated with QoC (Appendix C, Figure 9.6). Also, doctors as compared to other types of providers appeared to be associated with better QoC, although there seemed to be a large variation in quality

among the doctors. It should be noted here that there are only two CHW supervisors in this sample with 20 patient observations (Appendix C, Figure 9.7). Health worker knowledge score and number of training days did not appear to be associated with quality of care (Appendix C, Figure 9.8).

Based on the scatter plots, it appeared that identified motivation had a positive association with QoC, perhaps somewhat curvilinear (Appendix C, Figure 9.9). As expected, it also seemed that amotivation had a negative association with QoC. There were less clear patterns of relationship between (i) QoC and external regulation, and (ii) QoC and introjected regulation. Given that the factor scores appeared to be highly linearly related (Appendix C, Figure 9.9), they could result in multicollinearity, which would make the regression model estimates unstable and lead to inflation of the standard errors [155]. The test for multicollinearity using 'vif' command in STATA resulted in high VIF and tolerance (1/VIF) values, where identified and introjected motivation types had the highest values.

Based on average score for facility quality of care, being a female patient seemed to be associated with slightly higher QoC. However, as Figure 9.10 (Appendix C) shows there was a large variation within facilities for both male and female patients in quality of care, which were not reflected in a simple box plot of cluster-level means.

5.4.3 Null and treatment only model

The results of the two null models are provided in Table 5.3. The first model is a three-level model that reflects the original structure of the data where facilities are level-3, health workers are level-2 and patients are level-1 observations. As the Table shows the proportion of total variation in QoC that is attributable to differences across facilities relative to differences across health workers and patients within facilities is 9.0% only. Proportion of total variation in the outcome that is attributable to differences across health workers is 46%. Together they account for 55% of the variation in QoC, which means that 45% of the variation is due to within facility and health worker differences.

Following this, two different two-level models were fit to the data. First facilities were treated as level-2 observations (not shown in here), next health workers were treated as level-2 observations while nesting within facilities was ignored (Table 5.3). Comparing via likelihood ratio test the three-level model to the two-level model with a random intercept for facility, there was sufficient evidence to indicate that the variance in QoC among health workers in the same facility was significantly different from zero. Comparing the three-level model to the two-level model with a random intercept for a health worker, there was insufficient evidence in favor of positive variance in QoC among health facilities. Thus, adding $\psi(3)$ did not result in improved model fit once the clustering at health worker level was taken into account. Given these results, for the subsequent analyses a two-level model was used with health workers at level-2 and patients at level-1.

The treatment variable (level-2) was added to the model. As Table 5.4 shows, among patients in control arm facilities QoC z-scores would fall within +/- 2*√0.57 of the mean, i.e. between -1.59 and 1.44. Comparing patients in treatment group to patients in the control the expected change in z-score for QoC was 0.23 (p-value<0.05). The decrease in level-2 variance provided further evidence that treatment explained some of the variation in QoC observed between health workers. However, a large portion of between health worker variance was left unexplained. Thus, models with other covariates that could explain further the observed variance in quality between as well as within health workers were explored.

5.4.4 Other explanatory variables

Additional explanatory variables were selected based on the results of the EDA described earlier. Adding a health worker type changed the magnitude of treatment coefficient very little (0.21 as compared to 0.23), indicating that there was no confounding by type of health worker of the treatment-quality of care association, which would be expected given randomization. However, type of health worker appeared to have statistically significant and strong association with QoC, controlling for treatment. Compared to doctors, other types of health workers (nurses, midwives and community midwives) were associated with a lower quality of care (p-value<0.05). The finding confirms the observation made during the EDA where cluster-level quality of care scores appeared to be observably higher among doctors as compared to other health worker types. Also, the health worker type appeared to explain part of the observed between health worker

variance. Adding this variable to the treatment only model resulted in decrease of the level-2 variance from 0.57 to 0.51.

Controlling for treatment, health worker gender did not seem to have a strong association with QoC as indicated by the small statistically insignificant coefficient. Also, there was no change in level-2 variance between the model with treatment only and a model containing gender, in addition to treatment, as a predictor. Having up-to-date salary payment appeared to be a strong and significant predictor of QoC, controlling for treatment. Comparing patients of health workers who reported timely payment of salaries to patients of health workers who reported delays in their salaries, the expected change in z-score for QoC was 0.49 (p-value<0.05), which is more than twice the size of the effect of treatment. There was practically no change in the treatment coefficient (0.24 as compared to 0.23). Timely payment of salaries also seemed to explain part of the observed level-2 variance in the outcome, which decreased from 0.57 from treatment-only model to 0.52. Timely payment of salaries may be an indicator of the overall quality of facility management. There was no statistically significant association between knowledge score and quality of care. These results were held with different methods of handling the missingness in knowledge test items (Table 5.5).

Controlling for treatment, it appeared that factor score for identified regulation (factor 1) was statistically significantly associated with quality of care (Table 5.4). The expected increase in QoC z-score was 0.14 per standard deviation increase in identified motivation factor score (p-value<0.05). The between health worker variance has decreased, but only

slightly, in this model as compared to the model with treatment only (0.57 as compared to 0.55). As expected, the coefficient on treatment remained unchanged, indicating that QoC-treatment association was not confounded by motivation. Holding treatment constant, amotivation had a small but significant negative association with QoC (β=-0.16, p-value<0.05). Given the high correlation between the two motivation factor scores, they became statistically not significant once both of them were added to the same model. Moreover, this larger model neither had a significantly better fit nor explained more variance in the outcome. There was no association between external regulation and quality of care, holding treatment constant. The same applied to introjected regulation.

Controlling for treatment, type of health facility, specifically comparing patients in SCs to patients in CHCs, seemed to be a strong and significant predictor of quality of care (β=0.55, p-value<0.05). There was also some decrease in ψ estimate (0.54), suggesting further that type of health facility explained some of the observed variance in quality of care. Unlike type of facility, management type did not appear to be a significant predictor of quality of care. There was statistically significant association between equipment functionality index and quality of care (β=0.23, p-value<0.05). Equipment functionality index also seemed to explain part of the observed level-2 variance in the outcome, which decreased from 0.57 from treatment-only model to 0.52.

A model with treatment, controlling for type of health facility, equipment functionality index, identified motivation, up-to-date salary payment, and health worker type is shown in Table 5.4. The variables on type of health facility were no longer significant in this

model. Also, the coefficient for identified motivation decreased and became marginally significant (p-value=0.05). Comparing the two-level null model with this model, the between health worker variance declined from 0.58 to 0.43. The level-1 variance remained at the same level (0.47). Together these covariates explained 15% of the total variance in quality of care, while the proportion of level-2 variance explained by the covariates was 26%.

Next, level-1 covariates on patient age and gender were added separately to the model above. As expected there was small but significant effect of patient's age on quality of care. There was a negative association between a binary variable for age and quality of care z-score (β= -0.10, p-value<0.05), suggesting that controlling for other variables, health workers provided better care for those less than five years of age which was one of the target groups for PBP. There was also a small but significant association between gender and quality of care. Comparing male to female patients, the expected change in z-score for quality of care was 0.09 (p-value<0.05), holding treatment and other facility and health worker variables constant. Although based on the likelihood ratio test the model with patient level covariates fit the data significantly better than the model with only facility and health worker characteristics, the total variance in quality of care explained by the model remained at 15%.

The final model (Model L) showed that there was no confounding of the effect of treatment on quality of care by facility, health worker or patient characteristics (Table 5.4). Controlling for type of health facility, equipment functionality index, motivation,

up-to-date salary payment, health worker type, patient age and gender, comparing patients in treatment group to patients the control the expected change in z-score for quality of care was 0.22 (p-value<0.05). When comparing with the null model, the observed between health worker variance decreased from 0.58 to 0.43, while the within health worker variance remained virtually unchanged. A number of interaction terms were tested for evidence of treatment effect modification. However, there was no strong evidence that the effect of PBP was modified by facility or health worker characteristics.

Next, the full model (Model L) was tested for endogeneity at cluster-level that could be due to covariates having different within effects as compared to the between effects. The results of the Hausman's endogeneity test (Hausman's test statistic=11.4 with 2 d.f. signifying the number of covariates with both between- and within- cluster variation, p-value<0.05) indicated that there was evidence of misspecification of the model, specifically related to patient age and gender. However, as suggested by Rabe-Hesketh and Skrondal [159] if the estimated difference in the effects is small, using random-effects estimator would still be preferable to the fixed-effects estimator because overall, it has a smaller mean squared error. Histograms of predicted level-1 and level-2 residuals were examined for normal sampling distributions (Figure 5.4). While level-1 residuals from the full model appeared to have normal distribution, level-2 residuals appeared to have slight skew with thicker tails than what would be expected in normal distribution. Thus, the results were checked using the sandwich estimator that does not rely on correct model specification. The magnitude and significance of treatment coefficient remained unchanged (β=0.22, p-value<0.05).

5.4.5 Sensitivity analysis for the outcome variable

To ensure that association between treatment and quality of care was not sensitive to outliers, the model with treatment as the predictor of quality of care was re-fit after removing the 43 observations identified as severe outliers (raw quality of care score above 0.86 or z-score>2.91). As it could be expected, removing severe outliers, even if they made up only 2% of the sample, reduced the observed between- as well as within-health worker variance in quality of care. The level-2 variance in the two-level null model was reduced from 0.58 to 0.45, while the level-1 variance was reduced from 0.47 to 0.42. However, the magnitude of the treatment coefficient changed only slightly from 0.23 to 0.21. Moreover, it remained statistically significant at 5% level.

A final check was completed by removing moderate outliers as defined in Hamilton [103]: 220 out of 2,106 non-missing values (10.5%) were below (n=20) or above (n=200) the inner fences (raw quality of care score 0.11 – 0.66). After removing these observations from the dataset and re-fitting the null model, the level-2 variance was reduced to 0.20 while the level-1 variance was reduced to 0.22. It was clear that particularly high performers were removed from the sample where the highest quality of care score became 65.5%. Treatment coefficient became close to zero (β=.08, p-value>0.05). These results were determined to be biased, as the top 9.5% of observations were removed, making the mean quality of care score unreasonably low (36.3%) when compared to national and historical trends [40].

5.5 Discussion

PBP or supply-side incentive programs have been gaining popularity in many developed and developing countries over the past few years. There are 30 countries where RBF in health projects were recently completed, are currently implemented or under preparation [95]. While RBF projects include demand- and supply-side incentives, this provides some indication of the large amount of resources that are spent on improving health system performance through changes in provider incentives. Thus, the question of the effect of PBP on quality of care is of high policy relevance.

According to the results of an RBF evaluation in Rwanda [32], PBP are expected to have the largest effect on those services that need the least amount of effort from health workers but at the same time provide highest financial rewards. While providers in Afghanistan were not paid directly for aspects measured in the QoC index as defined in this study, the clinical processes of care that form the index can be viewed as services that require the least effort on the part of health workers. Also, health workers have more immediate control over clinical processes of care than quality of care measured through outcomes, including changes in health status, behavior and satisfaction of the population [174]. Based on the experience from Rwanda, it is expected that the most visible effect of RBF will be on services over which providers have the greatest control and which require the least effort [32]. The results of this study are consistent with this hypothesis.

The overall conclusion of this study is consistent with the results of the recent evaluation of the RBF Project in Afghanistan [75]. According to the results of the evaluation, PBP in

Afghanistan had a significant association with clinical processes of care, specifically those measured by the Client Background and Physical Assessment Index, Client Counseling Index, and Time Spent with Client. These indices are sub-components of the QoC index used in this study [75]. However, by using multi-level modeling methods, this study provides a deeper understanding of the effect of the PBP program in Afghanistan, complementing the evaluation. It examines heterogeneity in QoC between- and within- health workers and estimates the proportion of variance that is explained by the assigned treatment. By examining other variables that contributed to reduction in the observed variance in the outcome, this study will hopefully help to improve the design of the RBF program in Afghanistan as well as other countries. Specifically, as described below, it appears that investments in equipment as part of PBP may further improve clinical processes of care.

While PBP appeared to be a significant predictor of QoC, there was substantial heterogeneity between facilities and health workers that was unexplained by the treatment. This is an important finding because it shows that while programs such as PBP can lead to higher quality of care, there are important facility and health worker characteristics that need to be taken into account which can help explain variation in the observed outcome. In an attempt to do so, this study first conducted exploratory data analysis looking at the relationship between QoC index and variables such as facility type, infrastructure, management type, health worker gender, type of health worker, health worker motivation, health worker knowledge score, patient age, patient gender and other characteristics. Based on the results of this analysis, a series of multilevel models

were tested, revealing important variables that influence clinical processes of care, holding treatment constant. These models suggested that differences in clinical processes of care were attributable to many facility, health worker and patient characteristics. Thus, PBP is only one of many potential instruments for improving quality of care at the facilities [121], and it may not be the most effective approach as organizational aspects such as timeliness of salary payments can have an even larger effect on quality.

The findings of the study also point to ways of improving the design of programs such as PBP. According to a study on PBP and quality improvement in the USA, "the mere creation of financial incentives will not close the quality chasm; organizations must concurrently develop supporting infrastructure and develop new capabilities for continuous quality improvement" [175]. Limited success of traditional PBP programs in the USA is attributed in part to the fact that these programs are rarely paired with other complementary quality improvement activities such as training, infrastructure investments etc [175]. The significant effect of equipment on quality of care, controlling for treatment assignment, suggests that in addition to providing financial incentives to individual health workers, the program should make investments in improving facility infrastructure. This is also supported by evidence from PBP programs in other developing countries. In Rwanda, where the PBP program seemed to be successful in improving utilization and quality of health services, on average 23% of funds earned through the program were allocated for overall facility-level investments such as infrastructure and medical supplies [32].

There was a significant difference in expected QoC between CHCs and SCs, holding treatment constant. SCs are the lowest level of BPHS facilities included in the program. They are typically located in more rural remote areas of the country as compared to higher level facilities. Thus, the observed difference in expected QoC between the CHCs and SCs may be indicative of other factors that play an important role in QoC. SCs are much smaller than either BHCs or CHCs, often staffed with only one health worker, although in theory they should have one male nurse and one female community midwife [32]. Interestingly, once health worker characteristics were added to the model, type of facility was no longer statistically significant, which may imply that the initially observed effect of CHCs on quality of care was largely due to the presence of staff with higher qualifications, specifically doctors. Still, these results point to the importance of further investments in health posts and SC, as they are the first point of contact with health services for most of the population.

Similar to the finding of this study, a previous study in Afghanistan [39] shows no association between health worker gender and observed quality of care. It differs from the findings of the study by Das and Hammer [37] that found a positive association between female providers and higher quality of care. This could be due to the fact that Das and Hammer were comparing quality of care among doctors and had a large sample of female providers, while in this study 88.7% of observations comes from male providers. Moreover, 88.5% of observations made among doctors were among male doctors and only 11.5% of observations were among female providers of the same type.

Doctors were associated with higher level of quality as compared to other types of providers, holding treatment and other variables constant. This may reflect higher levels of training and knowledge that doctors have, although the variable on the knowledge test score and number of technical training days in the past 12 months did not appear to be associated with quality of care. This finding of positive association between being a doctor and providing better quality of care is consistent with previous findings in Afghanistan [39].

The lack of association between health worker knowledge score and QoC index was surprising. According to Das and Hammer [37], differences in doctors' competence or differences in doctors' incentives account for the variance in quality of care. Comparing doctors in public sector, whose salaries were fixed, with those in the private sector, who worked on the basis of fee-for-service, in Delhi, Das and Hammer noted that doctors who knew more also did more, if they were provided with right incentives, mainly in the form of financial rewards. At the same time, they found that less qualified private doctors seemed to provide higher quality of care than their better qualified colleagues working in public facilities because, as the authors of the study explain, they had incentives to do so [37]. Thus, controlling for incentives through the treatment variable, health worker competence, as measured by the score in the knowledge test, was expected to have a positive association with QoC. This was not the case in this study. Partly, it may be due to differences in samples between this study and the study in India. Das and Hammer [37] examined differences in quality of care and competence among doctors only, whereas this study included a wide range of clinical staff types and included only 75 doctors. In

addition, however, this may be due to how competence was measured in this study as compared to studies by Leonard et al [176] and Das and Hammer [37], which used vignettes that closely reflected competencies that were assessed during direct patient-provider observations. In this study, the knowledge score was based on a general test covering a wide range of technical knowledge such as infection prevention at facility level, general knowledge related to HIV transmission, TB diagnosis and treatment, storage of vaccines, newborn care, and family planning methods.

This study found a small but significant association between motivation, specifically identified regulation, and QoC. According to studies of quality of care in [36, 37, 177], it appears that there is a large gap between health worker competence and performance. These studies suggest that there is a positive association between QoC and health worker motivation. In the overview of variations in QoC in five low income countries, Das and Gertler [177] found that decreased effort or motivation can offset additional competence. The study by Leonard and Masatu [77] that specifically examines intrinsic motivation among health workers in Tanzania concludes that faced with the same types of incentives, certain types of health workers continue providing higher quality of care than their peers because they are driven by intrinsic motivation. These types of health workers seem to provide high QoC in a variety of organizational settings, many of which do not have mechanisms or culture that rewards additional effort. Serneels et al. [178] found that intrinsic motivation was one of the key variables explaining heterogeneity in the willingness of medical students in Ethiopia to work in rural areas.

As described in the previous chapter, in Afghanistan there was no association between PBP and motivation factors, including identified regulation, when comparing health workers assigned to the treatment group and health workers assigned to the control group. Also, the current study did not find significant association between external motivation and observed QoC. At the same time, there was small but significant association between identified motivation and QoC after controlling for treatment, i.e. assignment to PBP program. Thus, it seems that similar to the results in Tanzania [77], on average intrinsically motivated health workers provide higher quality of care regardless of the presence of external financial incentives.

This suggests that interventions that focus on improving identified motivation need more attention, including changes in organizational factors that encourage staff to voice their opinions and participate in decision-making, provide opportunities for professional growth, ensure good communication between the management and staff, and promote fair policies. Horizontal enlargement, which involves expanding jobs so they include more activities and have task configurations that are interesting for employees, and vertical enlargement, which expands jobs to give employees greater say over their duties, are recommended as key policies in enhancing identified motivation [19]. However, as Gagne and Deci acknowledge, [19] there is a need for further research on what specific organizational factors promote identified motivation.

There are certain limitations in this study. Firstly, the study focused only on one of the three aspects of the quality assessment framework proposed by Donabedian [31].

Improvements in the health status and behavior of the population are the ultimate aims of the RBF project in Afghanistan and other countries. Whether changes in clinical processes of care translate into changes in utilization of health services or health status of the population depends on a large set of factors that are beyond the control of individual health workers. Thus, they are less amenable to interventions such as bonus payments to individual health workers, which is the essence of the intervention described here. Secondly, due to the design of the RBF intervention in Afghanistan that, unlike the program in Rwanda, did not provide additional funds to the control arm facilities, it is impossible to separate the effect of increased "salaries" due to bonuses from the effect of payments based on performance [32]. Thirdly, the remaining large variance in the QoC index between- as well as within-health workers suggests that there are other important factors that were not measured in the study. Das and Gertler [177] in their multi-country overview of the variations in quality of care found that patient characteristics such as socio-demographic status were important predictors of QoC. This study, however, had only two patient-level variables and did not account for important characteristics such as SES at patient level. Fourthly, due to the small number of health workers observed per facility it was difficult to estimate between-facility as compared to between-health-worker variance. Lastly, using factor scores and treating motivation factors as observed variables gives biased results that are typically attenuated. Treating motivation factors as latent variables and using structural equation modeling would provide stronger evidence on the association between motivation and quality of care.

According to two recent Cochrane reviews of the effect of PBP on quality of health care services [61, 62], there is limited evidence of the effect of PBP programs on quality of care. This is due to weaknesses in study designs of many of these interventions. Studies of PBP programs included in these reviews were often based on non-experimental designs and failed to account for potential confounders. This study is based on data from cluster randomized trial and has a large probability-based sample. It examined the effect of PBP and explored other potential covariates that may contribute to improved quality of care. It also tested for treatment effect modification by key facility and health worker characteristics, providing further understanding of the process through which PBP may impact QoC. Taking advantage of multi-level modeling methods, it explored the relative contribution of different predictors of QoC in explaining the observed variation between as well as within health workers. While assessment of quality of care encompasses three distinct categories of structure, process and outcome and the effectiveness of the PBP program will differ for each of them, this paper provides evidence on the positive and significant effect of the program on clinical processes of care.

Table 5.1 Missing responses for at least one question in the knowledge test by type of health worker and treatment group

Type of health worker	Treatment (n=1,143)	Control (n=1,037)	Total missing	% of Total
Doctor	5.8	0.0	66	3.0
Nurse	11.2	10.6	238	10.9
Midwife	0.0	1.8	19	0.9
Community midwife	1.0	0.0	11	0.5
CHW supervisor	0.0	0.0	0	0.0
Total for all types	**17.9**	**12.4**	**334**	**15.3**

Table 5.2 Selected characteristics by treatment group

	Treatment (n=1,143)	Control (n=1,037)	n	p-value
Structural/ fixed characteristics				
Facility level				
By type of facility				
Sub-Centers	26.9	21.0	322	
BHC	49.1	58.2	1317	0.50
CHC	24.0	20.8	541	
By managing agency				
MOPH, without support	5.9	2.9	122	
MOPH, with support	54.7	55.0	1315	0.55
NGO only	39.4	42.0	743	
Health worker level				
Gender (male)	87.4	90.1	1914	0.51
Health worker type				
Doctor	26.6	26.5	608	
Nurse	65.1	66.7	1380	
Midwife	2.7	4.4	92	0.69
Community midwife	4.9	1.7	80	
CHW Supervisor	0.7	0.8	20	
Patient level				
Gender (male)	43.1	47.5	979	0.09
Age (<5 years of age)	48.4	49.3	1064	0.24
Modifiable characteristics				
Facility level				
Equipment functionality	78.87 (18.60)	82.77 (10.73)	2180	0.45
Pharmaceuticals & vaccines availability	76.11 (17.22)	78.24 (16.35)	2180	0.45
Infrastructure index	51.14 (28.43)	51.79 (30.06)	2180	0.90
Health worker level				
Health worker motivation factor scores				
Identified regulation	-0.11 (0.71)	-0.14 (0.71)	2171	0.74
External regulation	-0.26 (0.80)	-0.23 (0.76)	2171	0.75
Introjected regulation	-0.27 (0.76)	-0.22 (0.74)	2171	0.70
Amotivation	0.00 (0.71)	0.00 (0.73)	2171	0.99
Salary payment up-to-date	61.2	64.5	1353	0.64
Salary increase in the past 12 months	34.3	29.4	774	0.46
Training outside of the facility in the past 12 months (days)	5.61 (5.90)	5.70 (6.06)	2150	0.91
Knowledge score	63.64 (11.79)	64.33 (9.78)	2180	0.66

Note: The overall sample size is 2,180 observations. Sample sizes by treatment arm are given in columns. Data are mean (SD) or proportions, unless otherwise indicated. Means and proportions are based on weighted data. Test statistics are adjusted for multi-stage stratified cluster sampling using Taylor-linearized variance estimation. Strata with single sampling unit centered at overall mean. Sample characteristics are shown in non-standardized unit.

Table 5.3 Null model for the Quality of Care Index (z-scores)

	Three-level model				Two-level model			
Parameter	**Est**	**p-value**	**95% CI**		**Est**	**p-value**	**95% CI**	
Fixed part								
β0 [_cons]	0.04	0.44	-0.06	0.14	0.05	0.32	-0.05	0.15
Random part								
ψ(3)	0.10		0.02	0.59				
ψ(2)	0.49		0.32	0.73	0.58		0.48	0.71
θ	0.47		0.44	0.51	0.47		0.44	0.51
Derived estimates								
rho (3)	0.09							
rho (2)	0.46				0.55			
Log likelihood	-2491.3				-2491.8			

Note: The results are based on 2,097 observations at level-1, 254 observations at level-2 and 232 clusters at level-3.

Table 5.4 Maximum likelihood estimates for Quality of Care Index (z-score)

	Model A				Model F				Model I				Model L			
Parameter	**Est**	**p-value**	**95% CI**		**Est**	**p-value**	**95% CI**		**Est**	**p-value**	**95% CI**		**Est**	**p-value**	**95% CI**	
Fixed part																
β0 [_cons]	-0.07	0.31	-0.22	0.07	-0.07	0.32	-0.22	0.07	-0.12	0.51	-0.48	0.24	-0.10	0.59	-0.46	0.26
β1 [treatment]	**0.23**	**0.02**	**0.03**	**0.43**	**0.22**	**0.03**	**0.03**	**0.41**	**0.22**	**0.01**	0.04	0.40	**0.22**	**0.01**	**0.05**	**0.40**
β2 [nurse]									-0.40	0.00	-0.61	-0.18	-0.41	0.00	-0.63	-0.19
β3 [midwife]									-0.50	0.04	-0.97	-0.02	-0.49	0.04	-0.97	-0.02
β4 [cm midwife]									-0.17	0.47	-0.63	0.29	-0.14	0.55	-0.60	0.32
β5 [chws]									-0.36	0.48	-1.35	0.63	-0.37	0.47	-1.35	0.62
β6 [hw male]																
β7 [salary up-to-date]									0.39	0.00	0.20	0.57	0.39	0.00	0.21	0.58
β8 [bhc]									0.09	0.51	-0.17	0.35	0.08	0.54	-0.18	0.35
β9 [chc]									0.11	0.51	-0.22	0.44	0.10	0.54	-0.22	0.43
β10 [moph_support]																
β11 [ngo]																
β12 [equipment]									0.14	0.01	0.04	0.24	0.14	0.01	0.04	0.24
β13 [knowledge]																
β14 [fscore1]					0.14	0.01	0.04	0.24	0.09	0.05	0.00	0.18	0.09	0.05	0.00	0.18
β15 [patient age 5+]													-0.10	0.00	-0.16	-0.04
β16 [patient male]													0.09	0.01	0.02	0.15
Random part																
ψ	0.57		0.47	0.70	0.55		0.45	0.67	0.43		0.35	0.53	0.43		0.35	0.53
θ	0.47		0.44	0.51	0.47		0.44	0.51	0.47		0.44	0.51	0.47		0.44	0.50
Derived estimates																
R-sq	0.01				0.03				0.15				0.15			
Log likelihood	-2489.3				-2485.4				-2458.1				-2447.9			

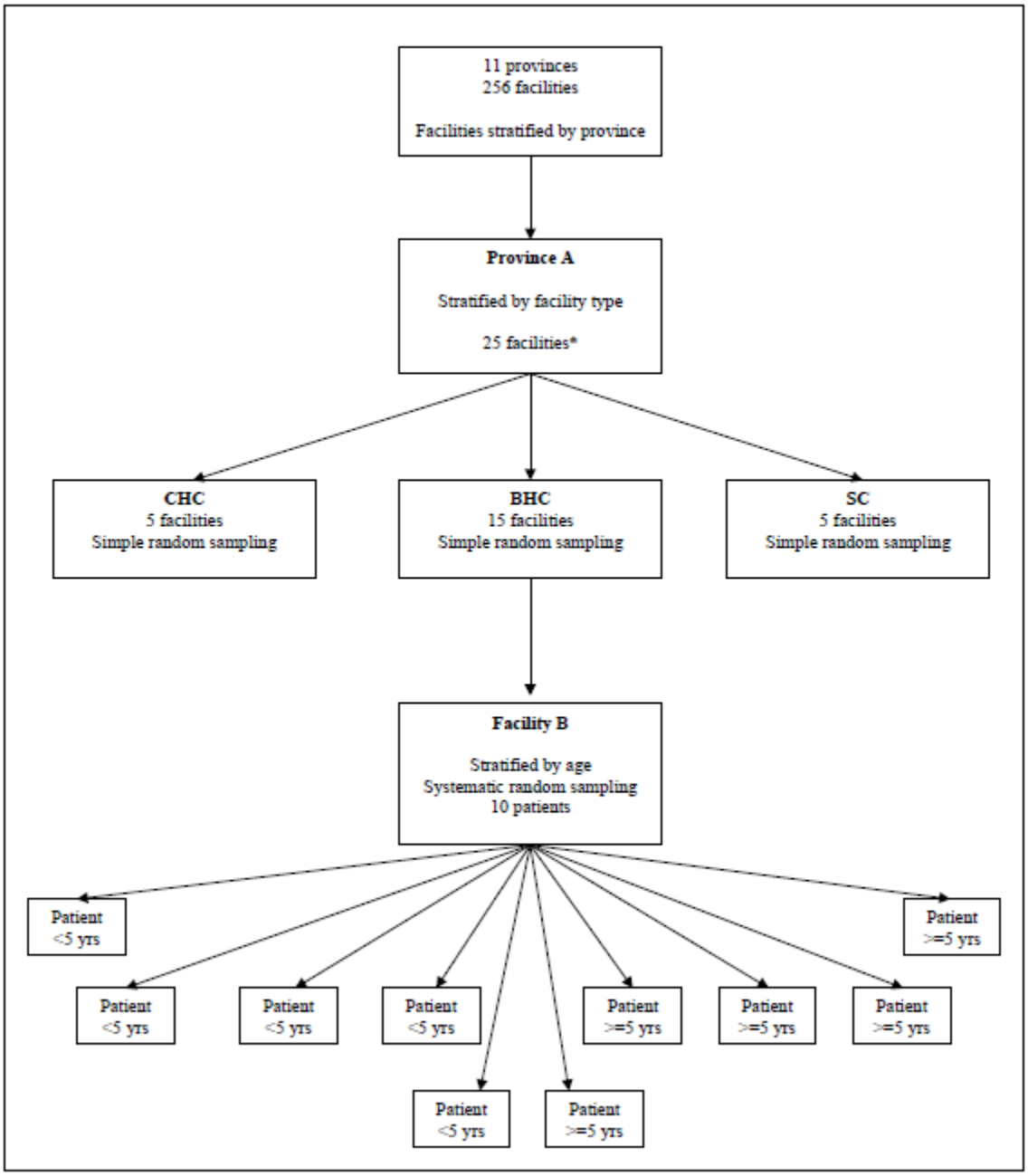
Note: Tables shows all the variables that were tested in smaller models. Model F is displayed as it is one of the main research questions. Model I is the model containing all covariates at cluster-level that were shown to be significant, prior to adding level-1 covariates. Model L is the final model.

The results are based on 2,097 observations at level-1 and 254 clusters. Given that all items on motivation score were missing for one health worker (one cluster, i.e. health worker and nine patients) were dropped from the original sample.
For health worker type, the reference category is doctors; for type of facility, the reference category is SC; for type of management, the reference category is MOPH only.
R-sq shows additional variance explained by the covariates in the model in comparison with the null model.

Table 5.5 Results for knowledge score effect with 3 different methods for missing data

	Complete case analysis (n=1,785)				Multiple imputation (n=2,106)				Missing responses assumed "Don't know"="Wrong"=0 (n=2,106)			
Parameter	**Est**	**p-value**	**95% CI**		**Est**	**p-value**	**95% CI**		**Est**	**p-value**	**95% CI**	
Fixed part												
β0 [_cons]	-0.08	0.29	-0.23	0.07	-0.07	0.36	-0.21	0.08	-0.07	0.38	-0.21	0.08
β1 [treatment]	**0.11**	**0.31**	**-0.10**	**0.31**	**0.22**	**0.03**	**0.03**	**0.42**	**0.23**	**0.02**	**0.03**	**0.42**
β13 [knowledge]	0.09	0.12	-0.02	0.21	0.04	0.33	-0.04	0.12	0.05	0.34	-0.05	0.15
Random part												
ψ	0.51		0.41	0.63	0.75		0.68	0.83	0.57		0.47	0.70
θ	0.47		0.44	0.50	0.69		0.67	0.71	0.48		0.45	0.51

Figure 5.1 NHSPA 2012-2013 sampling for patient observations



Note: * In provinces with less than 25 facilities, all facilities have been surveyed. In Kandahar, where only eight facilities were selected for RBF pilot, only eight facilities were included in RBF sample. In the following RBF pilot provinces – Parwan, Takhar, Kunduz & Saripul – more than 25 facilities per province were selected to ensure inclusion of matched pair facilities surveyed at baseline in 2010-2011.
For diagram displaying the sampling of health workers, see Chapter 4.

Figure 5.2 Quality of care by treatment arm

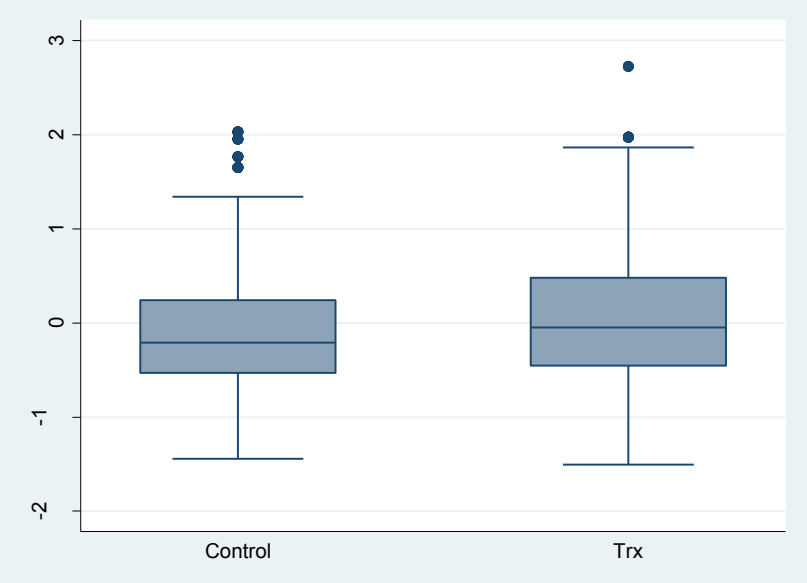


Figure 5.3 Check for normality of quality of care distribution by treatment arm

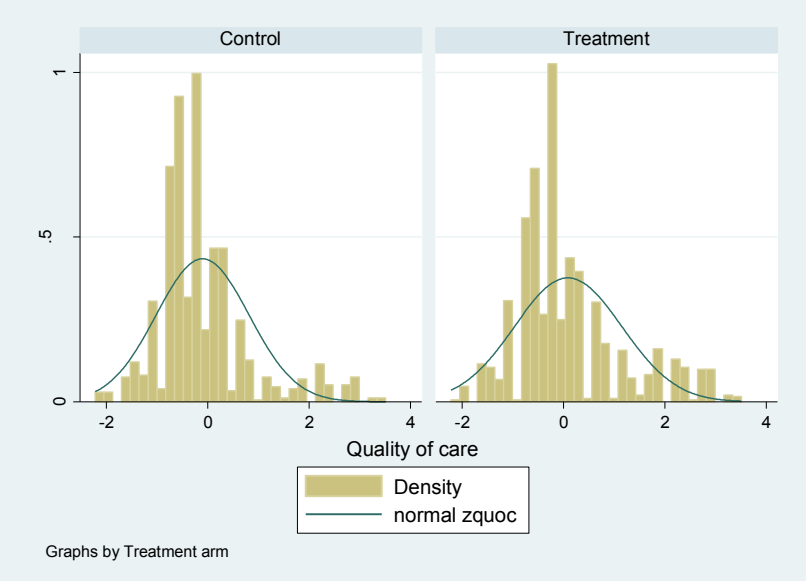
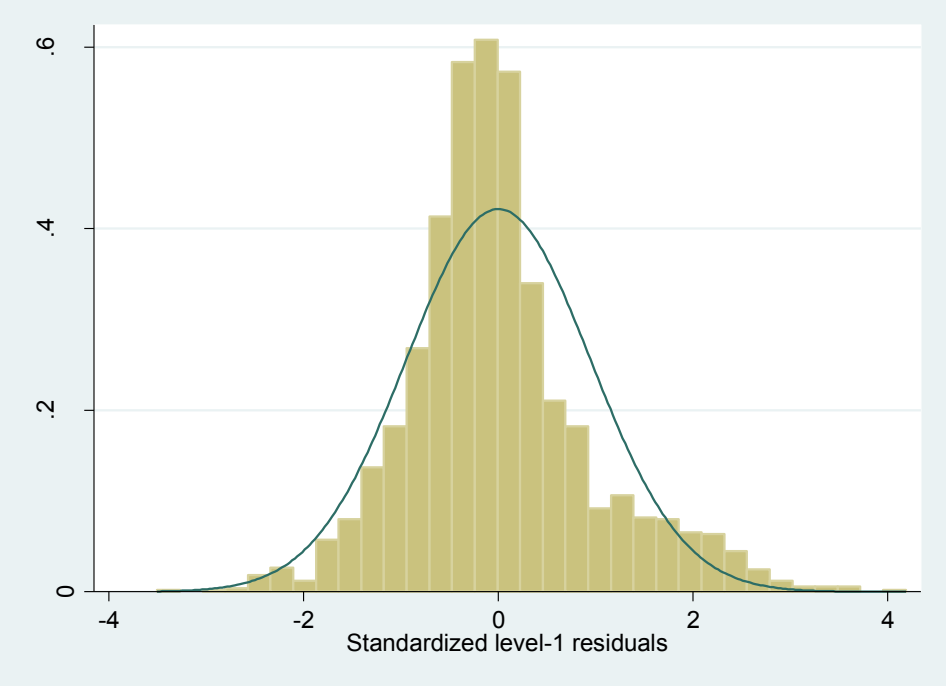
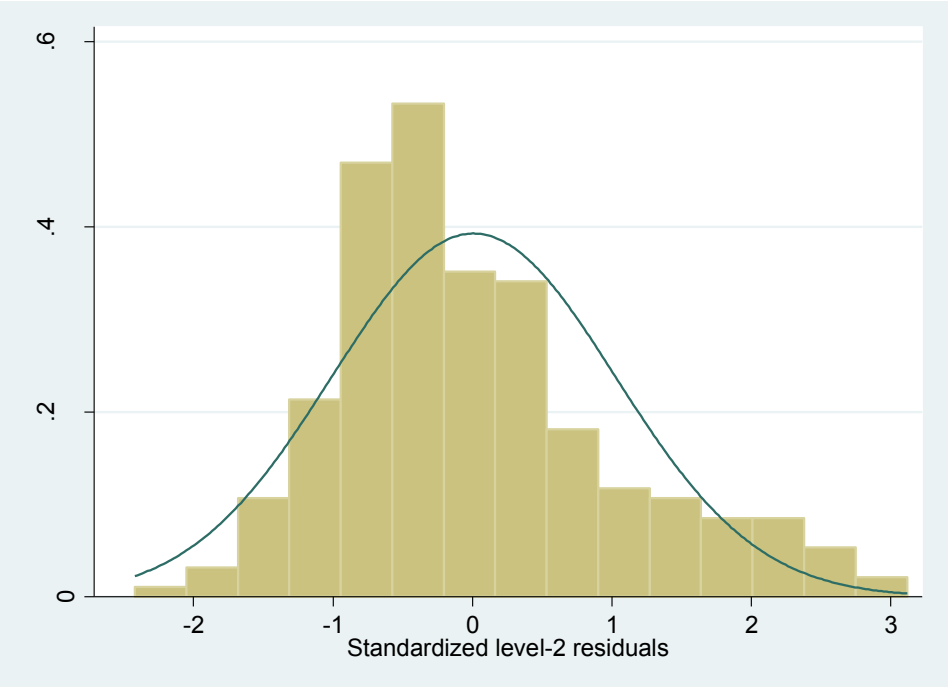


Figure 5.4 Residual diagnostics for Model L





6 Chapter 6: Conclusions

6.1 Summary of the findings

This thesis provides an in-depth analysis of the effects of performance-based payments (PBP) on health worker motivation and quality of care in Afghanistan. In Paper 1, a multidimensional work motivation scale based on a clear theoretical framework was developed and validated. The scale is short with only 15 items, but at the same time, provides a nuanced understanding of motivation that is important in designing human resource policies. The Dari-language scale demonstrated good psychometric properties. Moreover, the results of the measurement invariance test provided sufficient evidence that the scale could be applied to Pashto speakers as well. This is an important finding given that Pashto is the second official language after Dari, and Pashtuns make up the largest ethnic group in the country.

Paper 2 examined the effects of performance-based payments on different dimensions of health worker motivation. Based on the intention-to-treat analysis, it appeared that financial incentives in the form of performance-based payments did not have an effect on health worker motivation in Afghanistan. While other variables, such as gender and health worker type, were shown to be important predictors of different dimensions of motivation, the PBP intervention did not appear to have a significant association with any of the motivation factors. However, in cases with high proportions of non-compliers, i.e. those who did not receive the treatment, the ITT analysis results are often attenuated. Moreover, the results reflect the programmatic effect of the intervention, i.e. effect of the program on the outcome in realistic context where some participants receive treatment

while others do not. Efficacy of the intervention itself is more accurately assessed through methods such as complier-average causal effect (CACE). Given that a large proportion of health workers in this study reported that they did not receive pay-for-performance, it seemed reasonable to conduct analysis using the CACE method. The estimated complier-average causal effects for identified, external, and introjected motivation factors were negative (p-value<0.05). It appeared that PBP were negatively associated with motivation, with the largest negative effect observed for identified regulation.

Paper 3 examined the effect of performance-based payments program on quality of care defined as actions taken by the provider in making a diagnosis or treating the patient. It was measured using a standardized checklist that consisted of 16 items covering four areas of clinical quality of care: patient history, physical examination, patient counseling or communication, and time spent with patient. Based on the likelihood ratio test, it was determined that there was insufficient evidence in favor of a three-level model, and a two-level model with a random intercept for a health worker was fit to the data. The model with no covariates ("null model") showed that 55% of total variation in quality of care was attributable to differences across health workers, while 45% of the variation in quality of care was due to within health worker differences.

Comparing patients in the treatment group to patients in the control group, the expected change in z-score for quality of care was 0.23 (p-value<0.05). The decrease in level-2 variance provided further evidence that treatment explained some of the variation in the

outcome observed between health workers. Other covariates that were found to be significant predictors of quality of care were the following: equipment functionality index, motivation, up-to-date salary payment, and health worker type. Together, these covariates explained 15% of the total variance in quality of care, and 26% of the level-2 variance. In addition, patient age and gender had small but significant effects on quality of care. A number of interaction terms were tested for evidence of treatment effect modification. However, there was no strong evidence that the effect of PBP was modified by facility or health worker characteristics.

Together these three papers suggest that while PBP have a significant positive effect on measures of quality, this effect is not due to increased motivation. At the same time, holding treatment constant, it appeared that identified regulation (autonomous motivation type, close to intrinsic motivation) was statistically significantly associated with quality of (p-value<0.05). Thus, it appears that health workers who are more intrinsically motivated do provide higher quality of care, but their motivation is not affected by PBP. Moreover, while the treatment, i.e. PBP program, appeared to be a significant predictor of quality of care, it explained only a small portion of variation observed in quality of care between as well as within health workers. Health worker type, presence of functioning equipment, and timely payment of salaries seem to be more important predictors of quality of care, explaining a larger proportion of total variance in quality of care between health workers.

6.2 Study limitations and future research considerations

In addition to the specific limitations put forth in each paper, three general limitations apply to this dissertation. First of all, the study is based on a cross-sectional survey data thus limiting any causal inferences that can be made about the effects of PBP on motivation factors and quality of care. This also warns against drawing any definite conclusions regarding the effect of motivation on quality of care as the conceptual framework provided at the beginning of this study (Figure 1.3) shows how health worker performance feeds back to motivation via the consequences of worker performance. Thus, motivation at time t affects performance at time t, which has varying consequences at time t+1, including the amount of performance-based payments that a health worker receives for her efforts. This, in turn, has an effect on motivation at time t+1. As payments were made on a quarterly basis for the period of nearly two years, longitudinal data would have provided a more accurate understanding of the effect of performance-based payments on motivation and of motivation on quality of care.

Secondly, the aggregate data from MOPH shows there was a wide variation in the total amount of PBP earned by each facility. Some of the variation is due to the size of the facility. However, it is likely that, even among facilities of the same size as measured by the number of staff, there was a large variation in the amount of earned bonuses at the facility level. Moreover, it is likely that there was a large within facility variation of the amount that each health worker received. This study does not account for the variation in the amount that each facility and health worker received, and it is reasonable to assume

that they had an effect on both motivation and quality of care. It also does not account for the variation in intra-facility distribution types.

A more in-depth qualitative study would provide a richer understanding of the findings of this dissertation. The negative effect of PBP on external motivation found when comparing compliers in the treatment group, i.e. those who were assigned to PBP arm and reported to have actually received the payments, with compliers in the control group was surprising. Some explanation could be found in the data available on the design and implementation of the project through general MOPH documents and personal communication with the Ministry staff responsible for the project. However, a qualitative study including focus group discussions and interviews with NGOs running the facilities, the facility management, and health workers would provide a deeper understanding of reasons behind such surprising finding. The same applies to the findings regarding quality of care. Large unexplained variance observed between health workers as well as within health workers could be due to organizational structural and cultural factors that would be best explored through a qualitative study.

6.3 Policy implications

Access to health services depends to a large extent on availability of health workforce. On the one hand, there is evidence suggesting that low motivation contributes to shortages and low performance of individual health workers, weakening entire health systems. On the other hand, it seems that intrinsically motivated health workers can make a positive difference in access to and quality of health services even in settings with hard

resource-constraints. Health sector initiatives, such as Results-Based Financing in Health (RBF), are gaining popularity as governments and development partners try to address the complex challenges related to shortage of human resources in health, health worker performance and quality of health services. It is often assumed that improved financial incentives in the form of PBP improve health worker motivation and satisfaction. This, in turn, would lead to better health system performance, including quality of care. At the same time, there has been some difficulty in producing strong evidence on the impact of PBP programs on motivation and quality of care. There is even less evidence on the impact of motivation on health worker performance.

This is the first study on work motivation based on a large probability-based sample covering the entire country. It provides the first instrument on motivation that has been validated in Dari and is generalizable to Pashto language. Given the acute shortage and high turnover of health workers in Afghanistan and other countries in Central Asia, there is a high level of interest in improving health worker motivation among governments and international development partners in this region. An instrument that allows policy-makers to understand the nuanced multi-dimensional nature of motivation and to accurately measure it will hopefully contribute to the development of better informed policies based on evidence.

The study also showed that there are several design issues of the RBF program in Afghanistan that should be taken into account when designing the new phase of the project in Afghanistan or other RBF projects in the region. First of all, while enthusiasm

from health workers about the RBF project is welcome, it is necessary to ensure that there are no unrealistic expectations as to the size of the bonus payments. Secondly, while verification of data provided by facilities for the payment purposes maybe necessary, especially in countries with weakly developed health management information systems (HMIS), delays in payments must be avoided.

Thirdly, health workers must be made aware of the precise formula and mechanism of distribution of payments within their facilities. The experience of Burundi may be a useful start. Fourthly, while facility autonomy should be encouraged and it may seem reasonable to allow facilities to decide the mechanisms for within-facility distribution of PBP, some basic rules based on the experience and lessons learned so far appear to be necessary.

In Afghanistan, a large portion of facilities chose to distribute PBP earned by the facility in proportion to existing salaries, which meant that the most senior staff or staff with highest salary levels received the largest portion of PBP and not the staff most directly responsible for the achievement of the performance indicators, such as midwives and vaccinators. This is likely to have a negative effect on the external motivation of midwives and vaccinators. Also, in many developing country settings it is the "can do" component that is missing, not the "will do" component, as health workers are forced to provide care when they do not have basic infrastructure such as access to clean water, reliable source of electricity, or vital medical supplies and drugs. Thus, to increase the

effectiveness of PBP it seems necessary to ensure that part of the PBP earned by the facility is directed towards facility level expenditures.

Fourthly, as results of the third paper showed management issues such as timely payment of salaries and availability of functioning equipment are more important predictors of quality of care. While there is a widespread enthusiasm in international donor community about RBF in general and PBP specifically, the findings of this study imply that perhaps reforms targeting issues related to the overall public financial reforms and management of facilities may provide the same gains as PBP projects. This finding is especially important to consider when implementing PBP programs in countries with weak HMIS where large resources are spent on verification of data prior to making the payments, as the costs of the program may outweigh its benefits.

The third and fourth points above are related and indicate the ways in which PBP can be improved in the future. According to a study on PBP and quality improvement in the USA, "the mere creation of financial incentives will not close the quality chasm; organizations must concurrently develop supporting infrastructure and develop new capabilities for continuous improvement" [175]. Limited success of traditional PBP programs in the USA is attributed in part to the fact that these programs are rarely paired with other complementary quality improvement activities such as training, infrastructure investment etc [175].

The significant effect of presence of functioning equipment on quality of care, controlling for treatment assignment, suggests that in addition to providing financial incentives to individual health workers the program should make investments in improving facility infrastructure. This is also supported by the evidence from PBP programs in other countries, such as Rwanda. Implementation of quality improvement activities and changes in management practices alongside increased financial incentives to health providers are likely to increase the effect of PBP on quality of care.

7 Appendix A: Supplementary tables for Chapter 3

Table 7.1 Original 20-item scale

Please read carefully and circle one of the answers for each question, according to how you personally feel about the statement made. The answers range from 1 naan, which means you strongly disagree with the statement, to 4 naan, which means you strongly agree with the statement. There is no right or wrong answer, so please, do not spend long time thinking about each question, simply try to circle the answer that best describes your feelings.

4 NAAN = "STRONGLY AGREE" 3 NAAN = "AGREE"

2 NAAN = "DISAGREE" 1 NAAN = "STRONGLY DISAGREE"

*Example 1: The statement says, "**I really like winter**". So, if you like winter but it is not your favorite time of the year, you will choose 3 Naan. It means you agree with the statement but not that strongly, since you like it but not as much as maybe spring.*

	HOW MUCH DO YOU AGREE WITH THE STATEMENT?	**1 Naan Strongly Disagree**	**2 Naan Disagree**	**3 Naan Agree**	**4 Naan Strongly Agree**
250	I work in this job because it is part of the way in which I have chosen to live my life	1	2	3	4
251	I am glad that I work in this facility rather than other facilities in the country	1	2	3	4
252	I work here because it makes me feel important	1	2	3	4
253	I only work here so that I get paid at the end of the month	1	2	3	4
254	I frequently think of quitting this job	1	2	3	4
255	I feel I should personally take the credit or blame for the results of my work on this job	1	2	3	4
256	I do this job because my family would be disappointed if I quit	1	2	3	4
257	I work here because of opportunities for promotion	1	2	3	4
258	I sometimes feel my work here is meaningless	1	2	3	4
259	I work in this job because it allows me to decide how my work is organized	1	2	3	4
260	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)	1	2	3	4
261	I work in this job because it allows me to use my skills	1	2	3	4
262	I do this job because it gives me respect in the community	1	2	3	4
263	I work here because it is located in a safe area	1	2	3	4
264	I work here because of good benefits I receive (Note: all benefits – housing, transportation, anything else you receive – think overall)	1	2	3	4
265	It is hard for me to care very much about whether or not the work gets done right	1	2	3	4
266	I work in this job because I can accomplish something worthwhile in this job	1	2	3	4
267	I work here because it provides long term security for me	1	2	3	4
268	Since I've heard about opportunities to receive performance-based payments I've been working harder than before	1	2	3	4
269	I feel a very high degree of personal responsibility for the work I do on this job	1	2	3	4

Table 7.2 Guide for the focus group discussion

Guide to the focus group discussions
These questions are developed in order to understand the meaning of motivation and motivating factors for Afghan health workers. This is particularly important as Dari and Pashtu languages appear to often use the same word for motivation and encouragement.

The focus group discussions are to be conducted with two separate groups of 7-10 participants representing different types of EPHS (hospital) health workers.

1. What are things that come to your mind when someone says motivation? NOTE: Just let people shout out nouns, verbs, phrases and record them. Do not interrupt or correct.
2. What does being motivated mean to you?
3. Can you give an example of the situation when you felt motivated at work?
4. List aspects/factors that encourage a person to do his/her work well, in general, not specific to you or this work.
5. Why do you work here? NOTE: Ask both (4) and (5) because that way we can see if they understand (5) in the same way we do.
6. What made you choose to become a nurse/doctor?
7. Put in order of PRIORITY factors (so, have to RANK them) that affect your decision to work in your current job.
8. Which of these factors would make you want to change the job?

Factors that may affect your decision to work in your current job

Original ordered list	
Healing patients	
Salary	
Recognition and respect by community	
Recognition by the family	
Supportive supervisor	
Training opportunities	
Doing something positive for the country	
Future promotion	
Desire to do something useful	
Guilt	
Co-workers	
Ability to make my own decisions	
Benefits	
Job security	
Sympathy for those who are sick	
Ability to use my skills	
No other choice, too late to change careers	

Table 7.3 Distribution of work motivation scale items and overall motivation

		Facilities outside of RBF pilot, group 1						
Item No.	Item	N	Strongly disagree 1 (%)	Disagree 2 (%)	Agree 3 (%)	Strongly agree 4 (%)	Mean score	SD
1	**I work in this job because I have a chance to help other people through my work**	**431**	**1.2**	**1.4**	**25.5**	**71.9**	**3.68**	**0.56**
2	I work in this facility because it plays an important role in the community	431	1.6	1.6	20.0	76.8	3.72	0.58
3	**I feel I should personally take the credit or blame for the results of my work on this job**	**429**	**1.2**	**2.8**	**35.0**	**61.1**	**3.56**	**0.61**
4	I work in this job because it allows me to decide how my work is organized	431	1.6	3.5	45.7	49.2	3.42	0.64
5	**I work in this job because it allows me to use my skills**	**431**	**1.9**	**2.3**	**40.8**	**55.0**	**3.49**	**0.64**
6	I work in this job because I can accomplish something worthwhile in this job	431	0.5	0.9	20.7	78.0	3.76	0.48
7	**I feel a very high degree of personal responsibility for the work I do on this job**	**430**	**0.2**	**0.2**	**15.8**	**83.7**	**3.83**	**0.40**
8	**I work here because it makes me feel important**	**431**	**7.9**	**7.9**	**33.4**	**50.8**	**3.27**	**0.91**
9	**I do this job because my family would be disappointed if I quit**	**431**	**14.9**	**19.3**	**34.6**	**31.3**	**2.82**	**1.03**
10	**I do this job because it gives me respect in the community**	**431**	**9.5**	**14.4**	**34.3**	**41.8**	**3.08**	**0.97**
11	**I work in this job to gain God's grace**	**431**	**0.0**	**0.0**	**8.1**	**91.9**	**3.92**	**0.27**
12	I only work here to get so that I can get paid	431	28.1	27.6	30.2	14.2	2.30	1.03
13	**I work here because of opportunities for promotion**	**431**	**4.2**	**2.8**	**43.2**	**49.9**	**3.39**	**0.74**
14	**I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)**	**430**	**6.7**	**17.7**	**40.2**	**35.4**	**3.04**	**0.89**
15	**I work here because it is located in a safe area**	**431**	**6.3**	**16.0**	**33.9**	**43.9**	**3.15**	**0.91**
16	I work here because of good benefits I receive (Note: all benefits – housing, transportation, anything else you receive – think overall)	430	30.7	36.7	24.9	7.7	2.10	0.93
17	**I work here because it provides long term security for me**	**431**	**3.0**	**8.4**	**37.6**	**51.0**	**3.37**	**0.76**
18	**I frequently think of quitting this job**	**430**	**53.0**	**28.6**	**12.1**	**6.3**	**1.72**	**0.91**
19	**I sometimes feel my work here is meaningless**	**429**	**68.1**	**20.1**	**6.3**	**5.6**	**1.49**	**0.84**
20	**I don't care much about the quality of work here**	**431**	**68.9**	**19.3**	**6.3**	**5.6**	**1.48**	**0.84**
21	I work here because I have no other choice	429	11.0	15.4	36.6	37.1	3.00	0.98
	Overall, I feel very motivated to do my job	419	0.7	0.2	26.3	72.8	3.71	0.50

Facilities outside of RBF pilot, group 2							Facilities in RBF pilot						
N	Strongly disagree 1 (%)	Disagree 2 (%)	Agree 3 (%)	Strongly agree 4 (%)	Mean score	SD	N	Strongly disagree 1 (%)	Disagree 2 (%)	Agree 3 (%)	Strongly agree 4 (%)	Mean score	SD
439	**1.8**	**1.6**	**24.6**	**72.0**	**3.67**	**0.60**	**804**	**3.1**	**3.2**	**26.9**	**66.8**	**3.57**	**0.70**
439	0.9	2.3	21.2	75.6	3.72	0.55	803	1.4	4.6	26.5	67.5	3.60	0.64
438	**0.7**	**3.2**	**34.3**	**61.9**	**3.57**	**0.59**	**802**	**3.1**	**4.7**	**36.8**	**55.4**	**3.44**	**0.73**
438	0.5	4.1	45.7	49.8	3.45	0.60	804	2.1	5.4	46.8	45.8	3.36	0.68
439	**0.9**	**2.5**	**35.1**	**61.5**	**3.57**	**0.59**	**802**	**2.0**	**6.0**	**41.7**	**50.4**	**3.40**	**0.69**
438	0.2	0.7	21.5	77.6	3.76	0.46	801	1.1	2.3	24.3	72.3	3.68	0.58
438	**0.0**	**0.5**	**15.3**	**84.3**	**3.84**	**0.38**	**804**	**0.6**	**1.1**	**16.7**	**81.6**	**3.79**	**0.47**
439	**5.5**	**9.3**	**35.8**	**49.4**	**3.29**	**0.85**	**804**	**9.0**	**12.1**	**32.3**	**46.6**	**3.17**	**0.96**
439	**13.2**	**21.2**	**35.8**	**29.8**	**2.82**	**1.00**	**803**	**19.6**	**22.0**	**31.4**	**27.0**	**2.66**	**1.08**
439	**6.8**	**14.1**	**38.3**	**40.8**	**3.13**	**0.90**	**804**	**14.6**	**17.3**	**31.6**	**36.6**	**2.90**	**1.05**
438	**0.0**	**0.7**	**10.7**	**88.6**	**3.88**	**0.35**	**803**	**0.3**	**0.5**	**8.3**	**90.9**	**3.90**	**0.34**
439	23.9	32.4	31.4	12.3	2.32	0.97	802	28.6	30.6	29.4	11.5	2.24	0.99
439	**2.3**	**5.0**	**42.8**	**49.9**	**3.40**	**0.69**	**804**	**3.9**	**7.7**	**42.3**	**46.1**	**3.31**	**0.77**
439	**6.4**	**19.8**	**40.8**	**33.0**	**3.00**	**0.89**	**804**	**10.8**	**20.5**	**37.3**	**31.3**	**2.89**	**0.97**
439	**5.9**	**13.7**	**35.1**	**45.3**	**3.20**	**0.89**	**804**	**6.6**	**15.1**	**33.7**	**44.7**	**3.16**	**0.91**
439	28.0	40.3	24.2	7.5	2.11	0.90	803	35.2	32.5	25.3	7.0	2.04	0.94
438	**3.2**	**5.9**	**41.1**	**49.8**	**3.37**	**0.74**	**803**	**6.6**	**11.6**	**38.2**	**43.6**	**3.19**	**0.88**
439	**49.0**	**28.9**	**16.4**	**5.7**	**1.79**	**0.92**	**799**	**41.9**	**30.2**	**18.7**	**9.3**	**1.95**	**0.99**
439	**65.2**	**23.9**	**5.2**	**5.7**	**1.51**	**0.84**	**804**	**56.5**	**22.1**	**11.0**	**10.5**	**1.75**	**1.02**
438	**69.9**	**19.4**	**8.2**	**2.5**	**1.43**	**0.75**	**802**	**65.6**	**18.0**	**9.5**	**7.0**	**1.58**	**0.92**
438	9.4	15.5	36.8	38.4	3.04	0.96	802	12.7	18.6	35.2	33.5	2.90	1.01
419	0.0	0.2	26.0	73.8	3.73	0.46	797	0.9	1.3	24.7	73.2	3.70	0.54

8 Appendix B: Supplementary tables and graphs for Chapter 4

Table 8.1 Distribution of motivation scale items by treatment arm

		Treatment (n=430)				Control (n=375)					
		Strongly disagree	Disagree	Agree	Strongly agree	Strongly disagree	Disagree	Agree	Strongly agree	n	p-value
Identified regulation											
intr1	I work in this job because I have a chance to help other people through my work	4.6	2.5	28.2	64.6	2.8	4.5	21.3	71.4	804	0.15
intr3	I feel I should personally take the credit or blame for the results of my work on this job	3.3	3.5	36.7	56.6	3.8	5.0	37.6	53.6	802	0.77
intr7	I feel a very high degree of personal responsibility for the work I do on this job	0.2	1.5	18.0	80.3	1.2	0.9	16.9	81.0	804	0.39
intrj4	I work in this job to gain God's grace	0.2	0.6	9.1	90.1	0.2	0.3	7.3	92.3	803	0.76
External regulation											
extrg2	I work here because of opportunities for promotion	4.0	10.8	42.4	42.8	4.7	4.9	37.8	52.6	804	0.03
extrg3	I work in this facility because it has sufficient resources I need to do my job (medicine, equipment, infrastructure)	8.7	23.6	37.6	30.2	13.7	17.1	38.6	30.6	804	0.12
extrg4	I work here because it is located in a safe area	5.7	14.0	34.4	45.9	6.4	16.8	33.8	43.0	804	0.80
extrg6	I work here because it provides long term security for me	4.0	14.0	35.9	46.0	9.3	10.1	39.7	40.9	803	0.02
intr5	I work in this job because it allows me to use my skills	1.1	6.8	44.0	48.1	3.8	4.9	34.3	57.0	802	0.02
Introjected regulation											
intrj1	I work here because it makes me feel important	10.2	14.7	33.0	42.0	8.5	10.1	33.3	48.1	804	0.24
intrj2	I do this job because my family would be disappointed if I quit	20.1	23.9	29.0	27.0	19.6	21.6	30.7	28.1	803	0.91
intrj3	I do this job because it gives me respect in the community	13.4	19.8	33.1	33.8	16.1	17.3	30.2	36.4	804	0.63
Amotivation											
amot1	I frequently think of quitting this job	43.0	32.8	16.9	7.3	44.7	26.1	18.3	10.9	799	0.26
amot2	I sometimes feel my work here is meaningless	56.5	26.1	10.3	7.2	57.6	18.0	10.7	13.8	804	0.03
amot3	I don't care much about the quality of work here	62.1	21.9	9.7	6.3	68.5	14.8	8.9	7.8	802	0.18

Note: The overall sample size of the study is 805 observations. However, one observation had missing values on all of the work motivation scale items.

Table 8.2 Predictors of compliance status, CACE (without OER)

	Beta	S.E.	OR	p-value
BHC facility	-0.18	0.36	0.84	0.62
CHC facility	0.45	0.38	1.56	0.23
MOPH with support	1.63	0.51	5.10	0.00
NGO only	1.86	0.53	6.39	0.00
Male	-0.07	0.24	0.93	0.77
Nurse	-0.13	0.29	0.88	0.65
Midwife	0.16	0.38	1.18	0.67
Community MW	0.09	0.36	1.09	0.80
Vaccinator	-0.06	0.29	0.94	0.83
CHW Supervisor	-0.05	0.35	0.95	0.88
Intercept C1	-1.84	0.63		0.00

Table 8.3 CACE model with pretreatment covariates (without OER)

AIC			25109	
BIC (sample-size adj)			25259	
Loglikelihood			-12455	
Entropy			0.89	
			Compliers	**Non-compliers**
Final class proportions based on estimated			0.44	0.56
	Identified	**External**	**Introjected**	**Amotivation**
Treatment				
Compliers	-2.02**	-1.09**	-0.85**	0.56**
Non-compliers	0.95**	0.72**	0.32	-0.52
Nurse	-0.16	0.22	0.23	0.17
Midwife	-0.68**	0.29	0.60**	0.59**
Community MW	0.10	0.46**	0.57**	-0.11
Vaccinator	-0.37	0.34**	0.46**	0.29
CHW Supervisor	-0.24	0.37**	0.53**	0.49**
Intercepts	2.65**	1.67**	1.21**	-1.14**

Note: NC – non-compliers is reference class, their intercepts are at 0.
** statistical significance at 5% level.

9 Appendix C: Supplementary tables and graphs for Chapter 5

Figure 9.1 Quality of care between- and within-facility variance

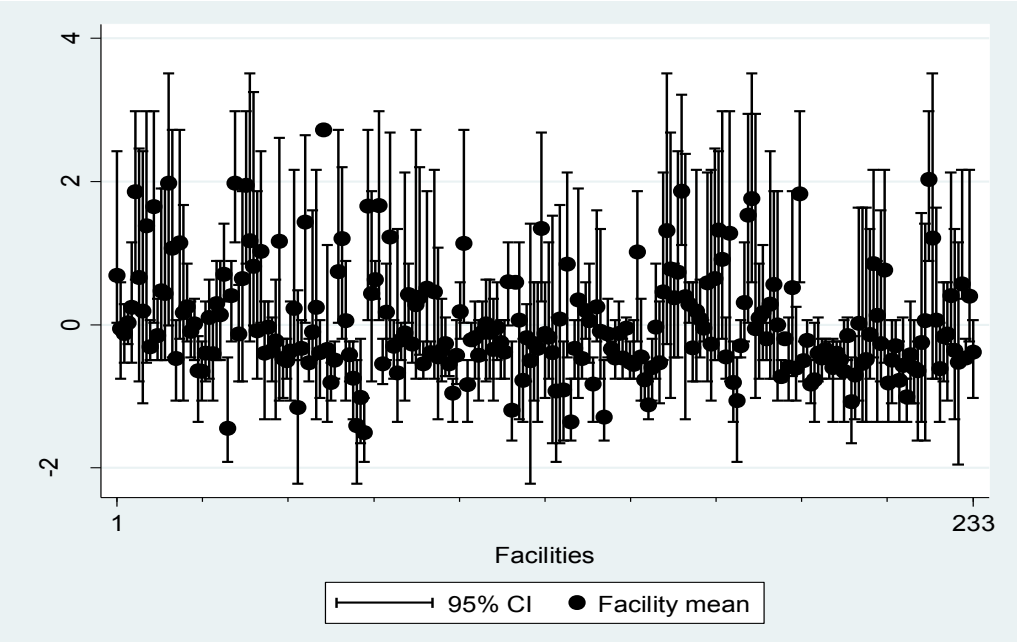


Figure 9.2 Quality of care observations by treatment arm

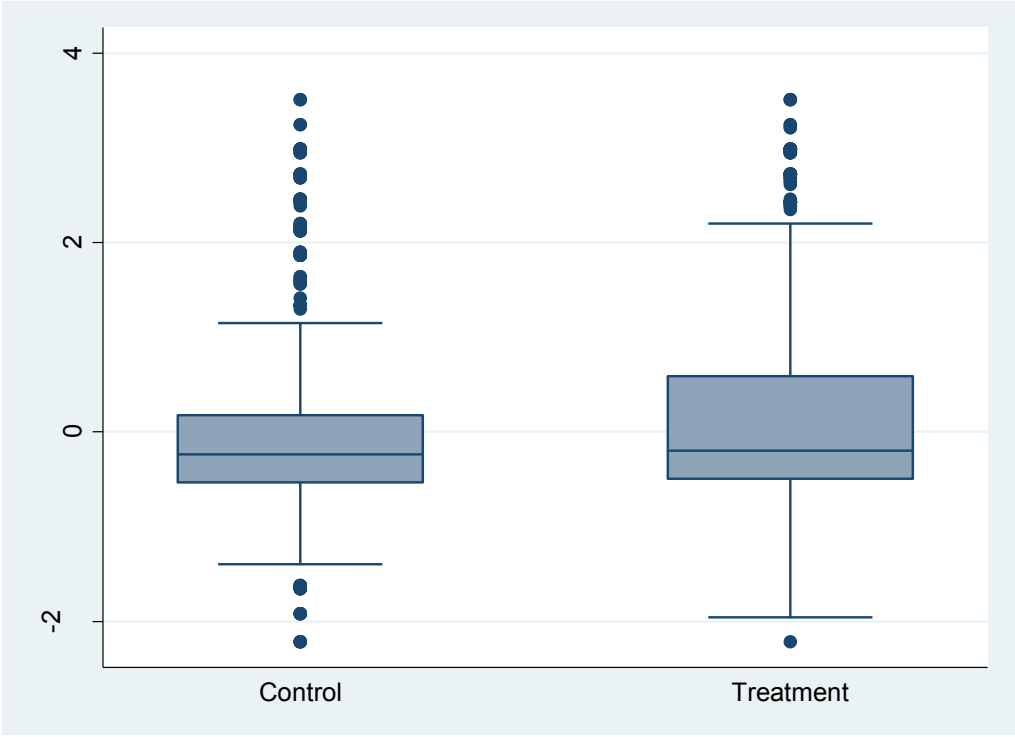


Figure 9.3 Quality of care by type of facility

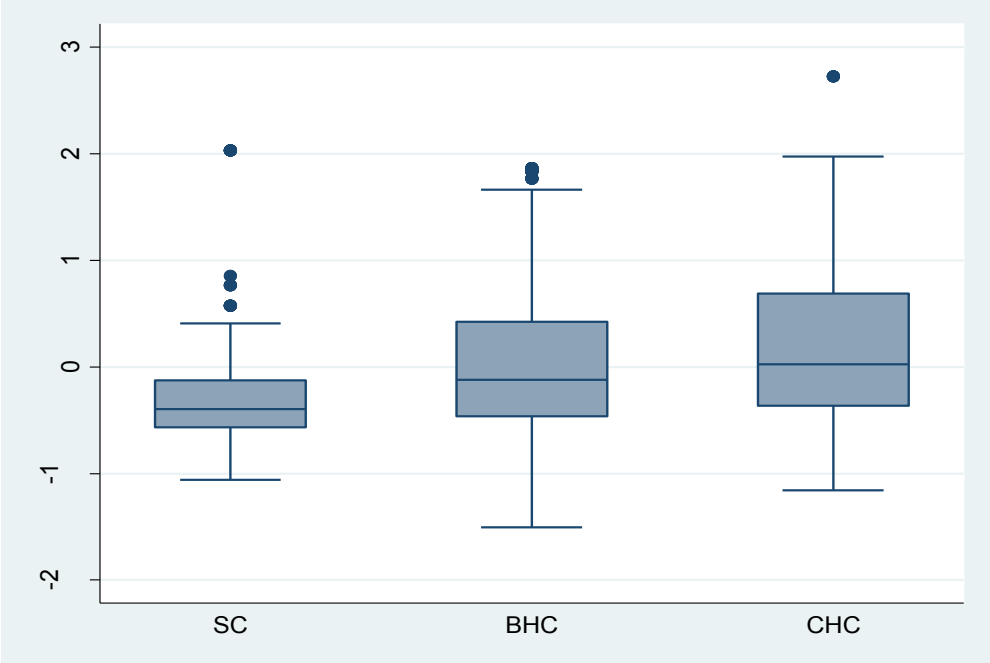


Figure 9.4 Quality of care by type of management

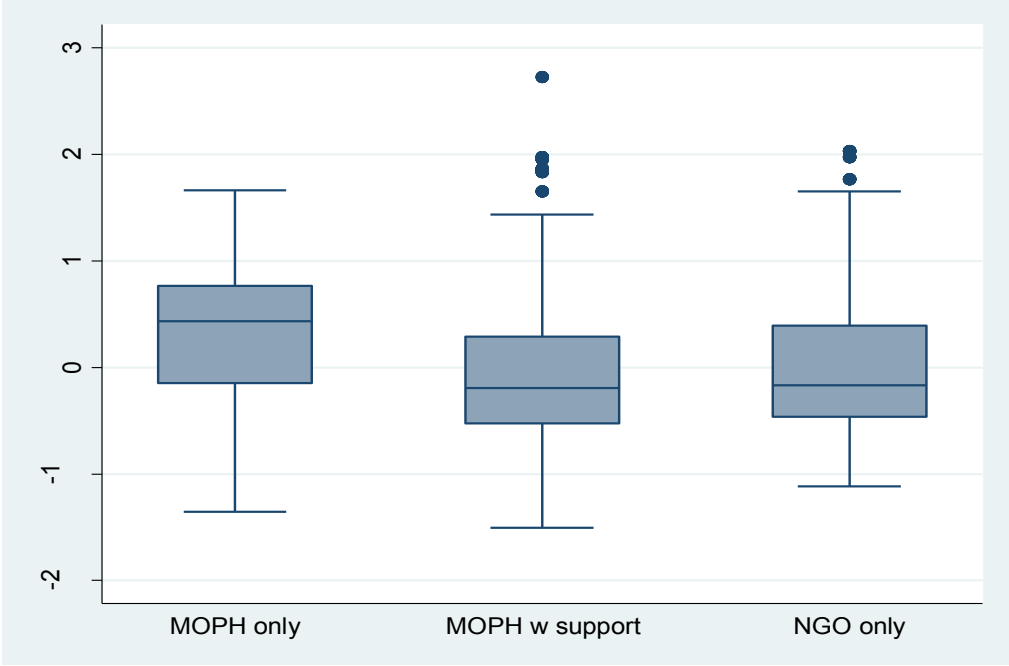


Figure 9.5 Quality of care, equipment functionality, pharmaceuticals and vaccines availability, and infrastructure

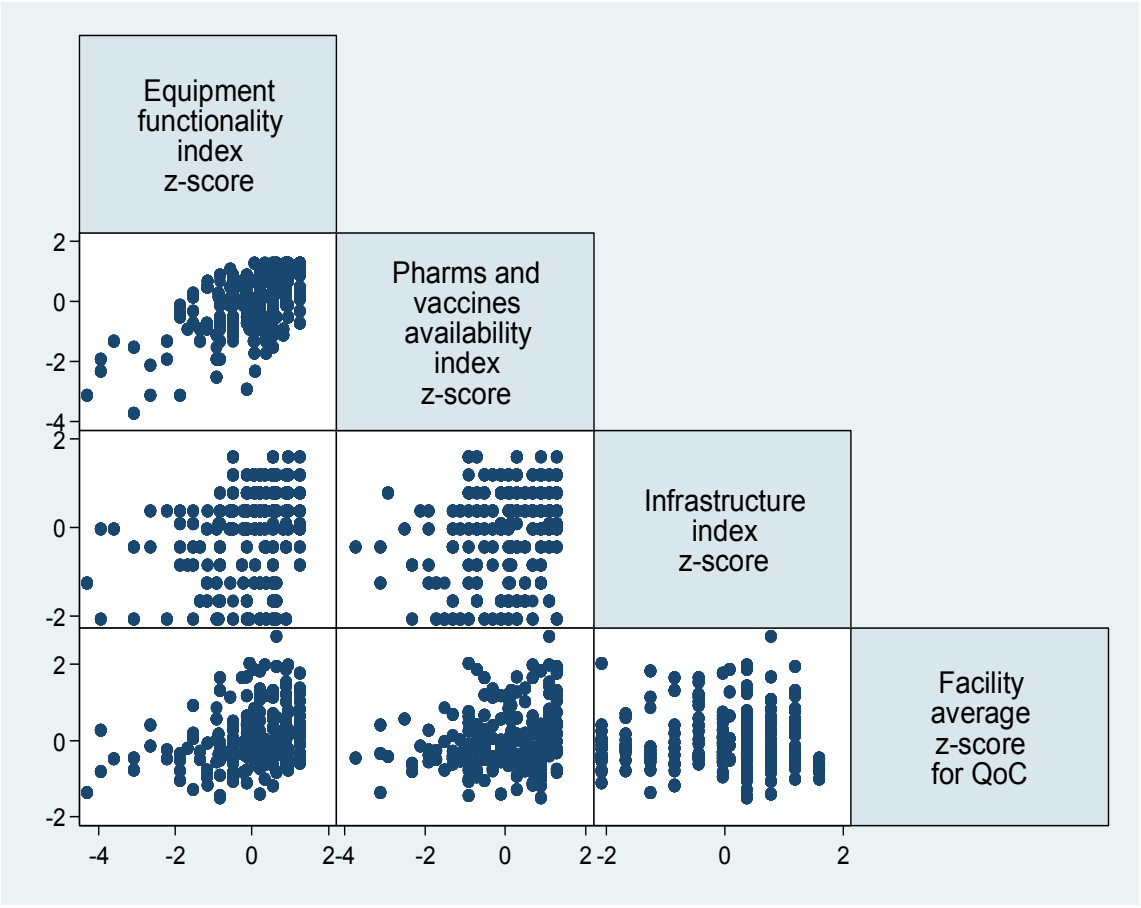


Figure 9.6 Quality of care by HW gender

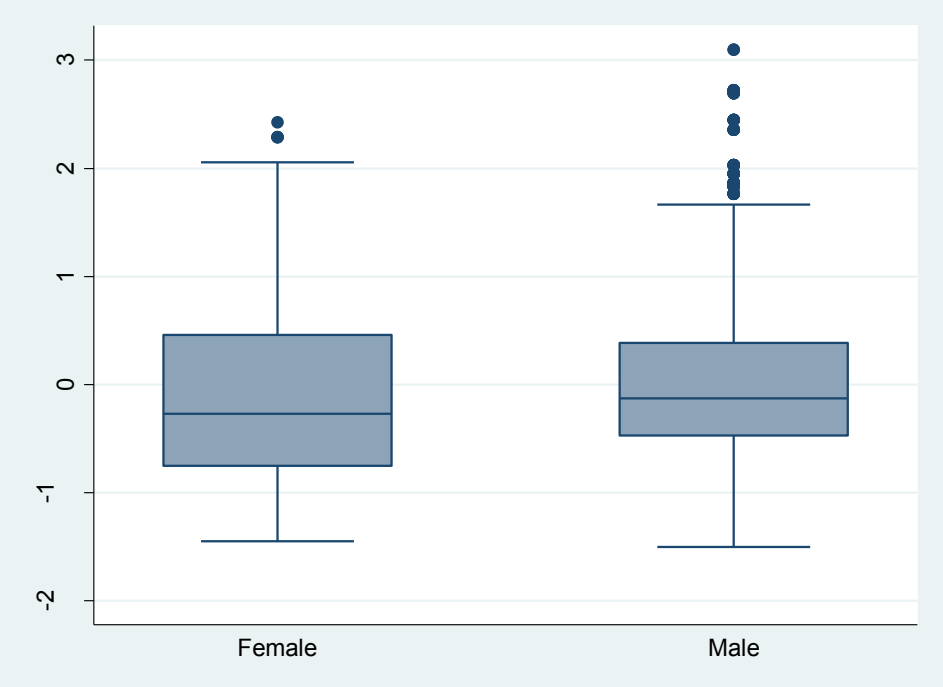


Figure 9.7 Quality of care by HW type

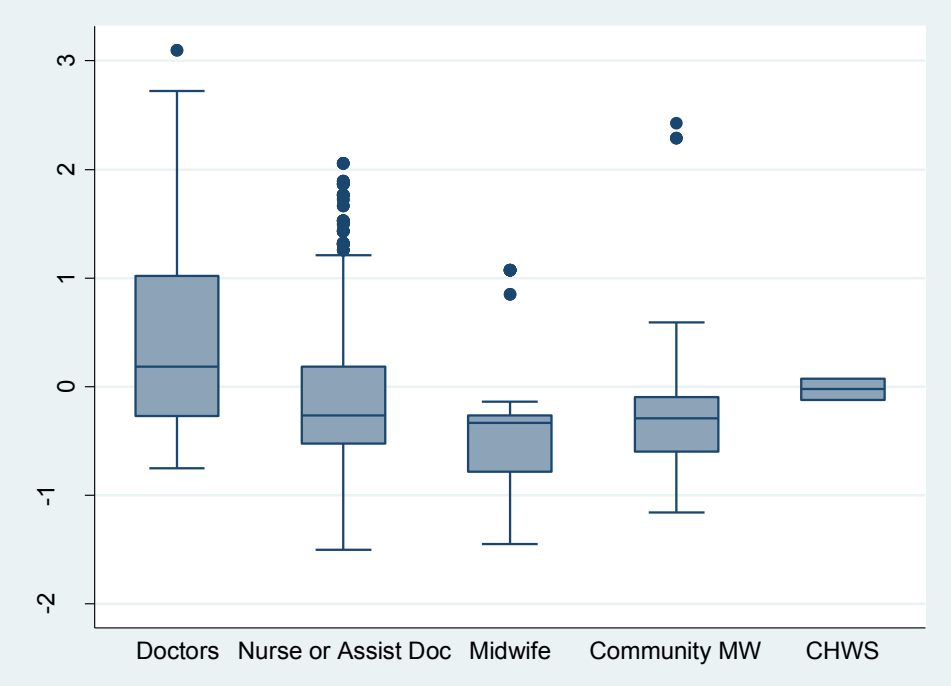


Figure 9.8 Quality of care, knowledge and training

Figure 9.9Motivation factor scores and quality of care

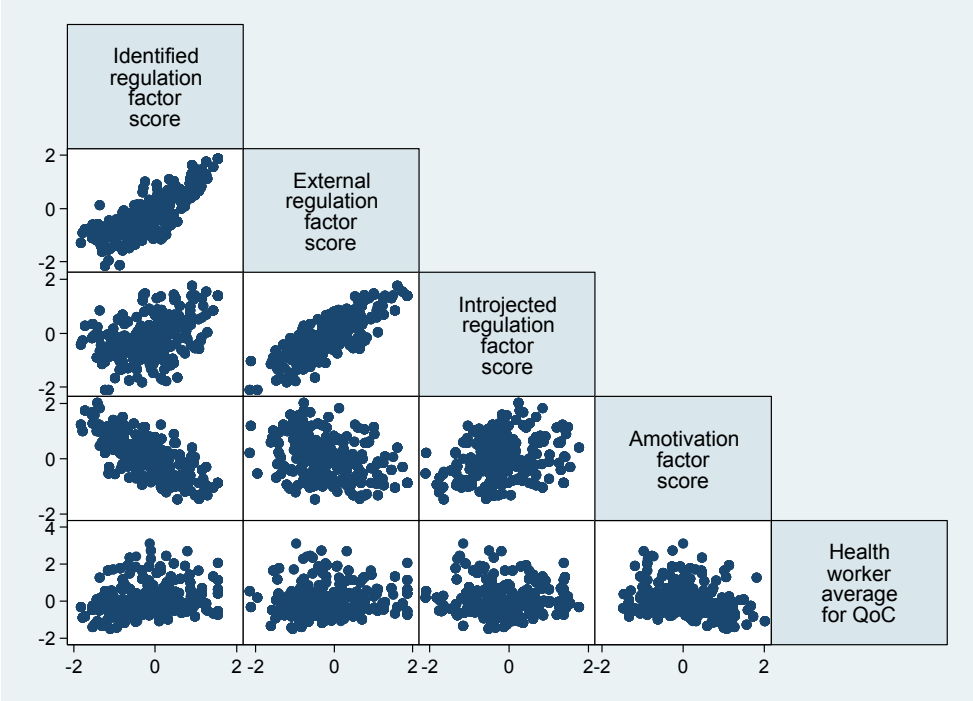
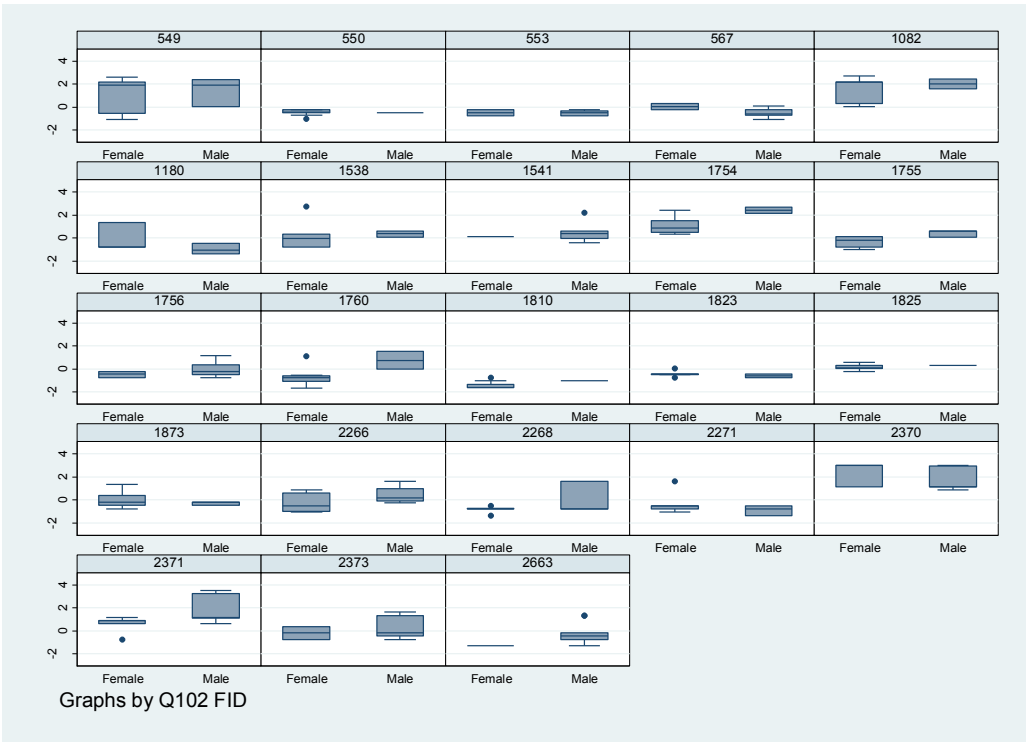
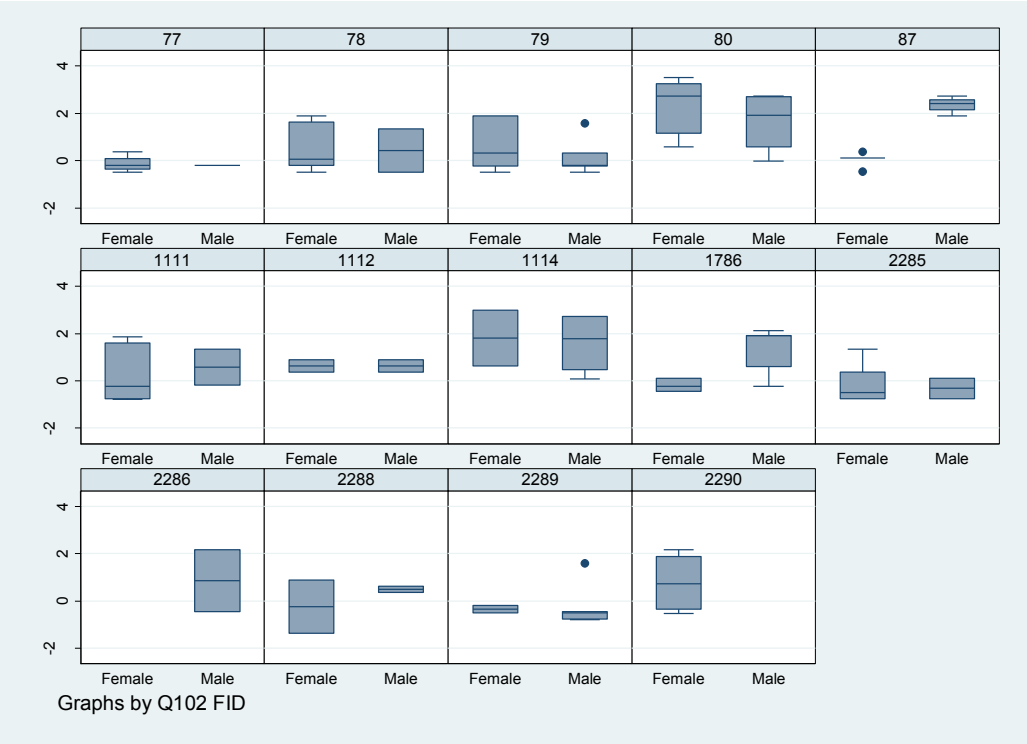


Figure 9.10 Quality of care by gender for selected facilities





References

1. **Afghanistan Overview** [http://www.worldbank.org/en/country/afghanistan/overview]
2. The World Bank: **World Development Indicators**. The World Bank; 2013.
3. Hogan MC, Foreman KJ, Naghavi M, Ahn SY, Wang M, Makela SM, Lopez AD, Lozano R, Murray CJL: **Maternal mortality for 181 countries, 1980–2008: a systematic analysis of progress towards Millennium Development Goal 5**. *The Lancet* 2010, **375**:1609-1623.
4. Alonge O: **The Effect of Contracting for Health Services on the Inequity of Utilization and Out-of-Pocket Health Expenditures in Rural Afghanistan**. *Dissertation.* Johns Hopkins University, International Health; 2012.
5. Ministry of Rural Rehabilitation and Development of the Islamic Republic of Afghanistan, Central Statistics Organization of the Islamic Republic of Afghanistan, European Commission: **Risk and Vulnerability Assessment 2007/8: A profile of Afghanistan**. Kabul; 2009.
6. The World Bank: **Strengthening Health Activities for the Rural Poor**. In *Project Appraisal Document*. Washington, DC: The World Bank; 2009.
7. United Nations Children's Fund: **Children in an Urban World**. In *The State of the World's Children 2012*. New York: UNICEF; 2012.
8. Meessen B, Soucat A, Sekabaraga C: **Performance-based financing: Just a donor fad or a catalyst towards comprehensive health-care reform?** *Bull World Health Organ* 2011, **89**(2):153-156.
9. Eichler R, Auxila P, Antoine U, Desmangles B: **Haiti: Going to Scale with a Performance Incentive Model**. In *Performance incentives for global health : potential and pitfalls* Edited by Eichler R, Levine R, the Performance-Based Incentives Working Group. Washington, DC: Center for Global Development 2009.
10. The World Bank: **Results-based financing intervention in BPHS facilities and hospitals in Afghanistan**. In *Concept Note*.
11. Shortell SM, Kaluzny AD: *Health Care Management: Organization Design and Behavior*. Fifth edition edition. Albany, NY: Delmar; 2006.
12. Kanfer R: **Measuring Health Worker Motivation in Developing Countries**. In *Major Applied Research 5 Working Paper 1*. Bethesda, MD: Partnerships for Health Reform Project, Abt Associates Inc; 1999.
13. Campbell JP, Pritchard R: **Motivation Theory in Industrial and Organizational Psychology**. In *Handbook of Industrial and Organizational Psychology* Edited by Dunnette M. Chicago: Rand McNally College Publishing Company; 1976:63-130.
14. Locke EA, Latham GP: **What should we do about motivation theory? Six recommendations for the twenty-first century**. *Acad Manag Rev* 2004, **29**:388–404.
15. Cummings TG, Bigelow J: **Satisfaction, job involvement, and intrinsic motivation: An extension of Lawler and Hall's factor analysis**. *Journal of Applied Psychology* 1976, **61**(4):523-525.

16. Peters D, Chakraborty S, Mahapatra P, Steinhardt L: **Job satisfaction and motivation of health workers in public and private sectors: cross-sectional analysis from two Indian states**. *Human Resources for Health* 2010, **8**(27).
17. Locke EA: **The Nature and Causes of Job Satisfaction**. In *Handbook of Industrial and Organizational Psychology*. Edited by Dunnette MD. Chicago: Rand McNally College Publishing Company; 1976.
18. Latham GP, Pinder CC: **Work motivation theory and research at the dawn of the twenty-first century**. *Annu Rev Psychol* 2005, **56**:485-516.
19. Gagne M, Deci EL: **Self-determination theory and work motivation**. *Journal of Organizational Behavior* 2005, **26**:331-362.
20. deCharms R: *Personal causation: The internal affective determinants of behavior*. New York, NY: Academic Press; 1968.
21. Deci EL, Ryan RM: **Self-determination research: Reflections and future directions**. In *Handbook of self-determination research*. Edited by Deci EL, Ryan RM. Rochester, NY: University of Rochester Press; 2002:431-441.
22. Levesque CS, Williams GC, Elliot D, Pickering MA, Bodenhamer B, Finley PJ: **Validating the theoretical structure of the Treatment Self-Regulation Questionnaire (TSRQ) across three different health behaviors**. *Health Education Research* 2007, **22**(5):691-702.
23. Tremblay MA, Blanchard CM, Taylor S, Pelletier LG, Villeneuve M: **Work Extrinsic and Intrinsic Motivation Scale: Its value for organizational psychology research**. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 2009, **41**(4):213-226.
24. Jones GR: *Organizational Theory, Design, and Change*. Fourth edition edition. Upper Saddle River, NJ: Prentice Hall; 2004.
25. Kao H, Sek-Hong N: **Work Motivation and Culture**. In *Motivation and Culture*. Edited by Munro D, Schmaker JF, Carr SC. New York, NY: Routledge; 1997.
26. Schwartz S: **Values and Cultures**. In *Motivation and Culture*. Edited by Munro D, Schmaker JF, Carr SC. New York, NY: Routledge; 1997.
27. **Self-Determination Theory: An Approach to Human Motivation & Personality** [http://www.selfdeterminationtheory.org/questionnaires]
28. Ryan RM, Connell JP: **Perceived locus of causality and internalization: Examining reasons for acting in two domains**. *Journal of Personality and Social Psychology* 1989, **57**:749-761.
29. Gagne M, Forest J, Gilbert M-H, Aube C, Morin E, Malorni A: **The Motivation at Work Scale: Validation Evidence in Two Languages**. *Educational and Psychological Measurement* 2010, **70**(4):628-646.
30. Institute of Medicine: **Crossing the Quality Chasm: A New Health System for the 21st Century**. Washington, DC: Institute of Medicine; 2001.
31. Donabedian A: **The Quality of Care. How Can It Be Assessed?** *The Journal of the American Medical Association* 1988, **260**(12):1743-1748.
32. Basinga P, Gertler PJ, Binagwaho A, Soucat ALB, Sturdy J, Vermeersch CMJ: **Eff ect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation**. *The Lancet* 2011, **377**(9775):1421-1428.

33. Franco LM, Franco C, Newton K, Nkhoma W: **Methods for assessing quality of provider performance in developing countries**. *International Journal for Quality of Health Care* 2002, **14**(1):17-24.
34. Onishi J, Gupta S, Peters DH: **Comparative analysis of exit interviews and direct clinical observations in Pediatric Ambulatory Care Services in Afghanistan**. *International Journal for Quality in Health Care* 2010:1-7.
35. Hansen PM, Peters DH, Edward A, Gupta S, Arur A, Niayesh H, Burnham G: **Determinants of primary care service quality in Afghanistan**. *International Journal for Quality in Health Care* 2008, **20**(6):375-383.
36. Leonard KL, Masatu MC: **Using the Hawthorne Effect to Examine the Gap Between a Doctor's Best Possible Practice and Actual Performance**. College Park, MD: Department of Agricultural and Resource Economics, University of Maryland; 2008.
37. Das J, Hammer J: **Money for Nothing: The Dire Straits of Medical Practice in Delhi, India**. In *World Bank Policy Research Working Paper 3669*. Washington, DC: The World Bank; 2005.
38. Hermida J, Nicholas DD, Blumenfeld SN: **Comparative validity of three methods for assessment of the quality of primary health care**. *International Journal for Quality in Health Care* 1999, **11**(5):429-433.
39. Hansen P: **Health Service Performance Assessment in Afghanistan**. *Dissertation.* Johns Hopkins University, International Health; 2007.
40. Johns Hopkins Bloomberg School of Public Health, Ministry of Public Health of the Islamic Republic of Afghanistan, Indian Institute of Health Management Research: **Afghanistan Basic Package of Health Services Balanced Scorecard: National Report 2012-13**. Kabul; 2013.
41. Eldridge C, Palmer N: **Performance-based payment: some reflections on the discourse, evidence and unanswered questions**. *Health Policy and Planning* 2009, **24**:160-166.
42. Musgrove P: **Rewards for Good Performance or Results: A Short Glossary**. Washington, DC: The World Bank; 2011.
43. Magrath P, Nichter M: **Paying for performance and the social relations of health care provision: An anthropological perspective**. *Social Science & Medicine* 2012, **75**(10):1778-1785.
44. Levacic R: **Teacher Incentives and Performance: An Application of Principal – Agent Theory**. *Oxford Development Studies* 2009, **37**(1):33-46.
45. Mathauer I, Imhoff I: **Health worker motivation in Africa: the role of non-financial incentives and human resource management tools**. *Human Resources for Health* 2006, **4**(24).
46. Amare Y: **Non-Financial Incentives for Voluntary Community Health Workers: A Qualitative Study**. In *Working Paper No 1* Addis Ababa: The Last Ten Kilometers Project; 2009.
47. Franco LM, Bennett S, Kanfer R, Stubblebine P: **Health Worker Motivation in Jordan and Georgia: A Synthesis of Results**. In *Major Applied Research 5, Technical Paper 3*. Bethesda, MD: Partnerships for Health Reform Project, Abt Associates Inc.; 2000.

48. Chandler CIR, Chonya S, Mtei F, Reyburn H, Whitty CJM: **Motivation, money and respect: A mixed-method study of Tanzanian non-physician clinicians**. *Social Science & Medicine* 2009, **68**(11):2078-2088.
49. Komaki JL, Coombs T, Schepman S: **Motivational implications of reinforcement theory**. In *Motivation and work behavior*. Fifth edition edition. Edited by Steers RM, Porter LW. New York, NY: McGraw Hill; 1991:87-105.
50. DeCharms R, Muir MS: **Motivation: Social Approaches**. *Annual Review of Psychology* 1978, **29**:91-113.
51. Deci EL, Ryan R: *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum; 1985.
52. Deci EL, Koestner R, Ryan RM: **A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation**. *Psychological Bulletin* 1999, **125**(6):627–668.
53. Lawler EE, Hackman R: **Impact of employee participation in the development of pay incentive plans: A field experiment**. *Journal of Applied Psychology* 1969, **53**(6):467-471.
54. Mitchell TR: **Matching motivational strategies with organizational contexts**. In *Research in Organizational Behavior. Volume 19*. Edited by Cummings LL, Staw BM. London: JAI Press Inc; 1997.
55. Lee TW, Locke EA, Phan SH: **Explaining the assigned goal-incentive interaction: the role of self-efficacy and personal goals**. *J Manag* 1997, **23**:541-559.
56. Koestner R, Ryan RM, Bernieri F, Holt K: **Setting limits on children's behavior: the differential effects of controlling vs. informational styles on intrinsic motivation and creativity**. *Journal of Personality* 1984, **52**:231–248.
57. Ryan RM: **Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory**. *Journal of Personality and Social Psychology* 1982, **43**:450–461.
58. Deci EL: **The effects of contingent and noncontingent rewards and controls on intrinsic motivation**. *Organizational behavior and human performance* 1972, **8**(2):217–229.
59. Locke EA: **Personnel attitudes and motivation**. *Annu Rev Psychol* 1975, **26**:457-480.
60. Rosenthal MB, Frank RG: **What Is the Empirical Basis for Paying for Quality in Health Care?** *Medical Care Research and Review* 2006, **63**(2):135-157.
61. Scott A, Sivey P, Ouakrim D, Willenberg L, Naccarella L, Furler J, Young D: **The effect of financial incentives on the quality of health care provided by primary care physicians**. *Cochrane Database of Systematic Reviews 2011* 2011(9).
62. Witter S, Fretheim A, Kessy FL, Lindahl AK: **Paying for performance to improve the delivery of health interventions in low- and middle-income countries** *Cochrane Database of Systematic Reviews 2012* 2013(2).
63. United Nations Children's Fund, Central Statistics Organization: **Transitional Islamic Government of Afghanistan. Multiple Indicator Cluster Survey – 2003, Afghanistan**. Kabul; 2003.

64. Arur A, Peters D, Hansen P, Mashkoor MA, Steinhardt LC, Burnham G: **Contracting for health and curative care use in Afghanistan between 2004 and 2005**. *Health Policy and Planning* 2010, **25**(2):135-144.
65. Belay TA (Ed.): *Building on early gains in Afghanistan's health, nutrition, and population sector: challenges and options*. Washington, DC: The World Bank; 2010.
66. Ministry of Public Health: **A Basic Package of Health Services for Afghanistan – 2010/1389**. Kabul: Islamic Republic of Afghanistan Ministry of Public Health; 2010.
67. Johns Hopkins Bloomberg School of Public Health, Indian Institute of Health Management Research: **Afghanistan Health Survey (AHS) 2012**. Kabul; 2013.
68. The World Bank: **Results-based financing intervention in BPHS facilities and hospitals in Afghanistan**. In *Operations Manual*. The World Bank; 2010.
69. Oshang N: **RBF - question**. Kabul; 2013.
70. Oshang N: **Questions on RBF** Kabul; 2011.
71. Oshang N: **Afghanistan RBF Scheme Incentive Distribution**. Kabul; 2012.
72. Tawfiq E: **Re: question on RBF mechanisms**. Edited by Dale E. Kabul; 2011.
73. Smith PG, Morrow RH (Eds.): *Field Trials of Health Interventions in Developing Countries: A Toolbox*. Second edition edition. London: Macmillan Education Ltd; 1996.
74. Ahmed S: **Methods for Program Evaluation: Design and Analysis** In *Methods in Analysis of Large Population Surveys*. Baltimore, MD: Johns Hopkins School of Public Health; 2011.
75. Johns Hopkins Bloomberg School of Public Health, Indian Institute of Health Management Research: **Impact Evaluation of the Results-Based Financing Intervention in Afghanistan**. Kabul; 2013.
76. Hayes RH, Moulton LH: *Cluster Randomized Trials*. Boca Raton: CRC Press; 2009.
77. Leonard KL, Masatu MC: **Professionalism and the Know-Do Gap: Exploring Intrinsic Motivation Among Health Workers in Tanzania**. *Health Economics* 2010, **19**:1461–1477.
78. Vujicic M, Ohiri K, Sparkes S: *Working in Health: Financing and Managing the Public Sector Health Workforce*. Washington, DC: The World Bank; 2009.
79. Dussault G: **Improving the Performance of the Health Workforce: From Advocacy to Action**. In *Priorities and Strategies in Human Resources for Health*. 47th Directive Council, PAHO/WHO; 2006.
80. World Health Organization, The World Bank: **Improving Health Workforce Performance. Issues for Discussion: Session 4**. In *High Level Forum on the Health Millenium Development Goals*. World Health Organization, The World Bank; 2003.
81. Willis-Shattuck M, Bidwell P, Thomas S, Wyness L, Blaauw D, Ditlopo P: **Motivation and retention of health workers in developing countries: a systematic review**. *BMC Health Services Research* 2008, **8**(1):247.
82. Serneels P, Montalvo JG, Pettersson G, Lievens T, Butera JD, Kidanu A: **Who wants to work in a rural health post? The role of intrinsic motivation, rural**

background and faith-based institutions in Ethiopia and Rwanda. *Bulletin of the World Health Organization* 2010, **88**(5):342-349.
83. Mutale W, Ayles H, Bond V, Mwanamwenge MT, Balabanova D: **Measuring health workers' motivation in rural health facilities: baseline results from three study districts in Zambia**. *Human Resources for Health* 2013, **11**(8).
84. Raykov T, Marcoulides GA: *Introduction to Psychometric Theory*. New York, NY: Routledge Taylor & Francis Group; 2011.
85. Prytherch H, Leshabari MT, Wiskow C, Aninanya GA, Kakoko DCV, Kagone M, Burghardt J, Kynast-Wolf G, Marx M, Sauerborn R: **The challenges of developing an instrument to assess health provider motivation at primary care level in rural Burkina Faso, Ghana and Tanzania**. *Glob Health Action* 2012, **5**.
86. Mbindyo PM, Blaauw D, Gilson L, English M: **Developing a tool to measure health worker motivation in district hospitals in Kenya**. *Human Resources for Health* 2009, **7**(1):40.
87. Rahman SM, Ali NA, Jennings L, Seraji MHR, Mannan I, Shah RZ, Mahmud ABA, Bari S, Hossain D, Das MK *et al*: **Factors affecting recruitment and retention of community health workers in a newborn care intervention in Bangladesh**. *Human Resources for Health* 2010, **8**(1):12.
88. Pillay R: **Work satisfaction of professional nurses in South Africa: a comparative analysis of the public and private sectors**. *Human Resources for Health* 2009, **7**(15).
89. Liu JA, Wang Q, Lu ZX: **Job satisfaction and its modeling among township health center employees: a quantitative study in poor rural China**. *BMC Health Services Research* 2010, **10**(1):115.
90. Prytherch H, Kagoné M, Aninanya GA, Williams JE, Kakoko DC, Leshabari MT, Yé M, Marx M, Sauerborn R: **Motivation and incentives of rural maternal and neonatal health care providers: a comparison of qualitative findings from Burkina Faso, Ghana and Tanzania**. *BMC Health Services Research* 2013, **13**(149).
91. Malik A, Yamamoto S, Souares A, Malik Z, Sauerborn R: **Motivational determinants among physicians in Lahore, Pakistan**. *BMC Health Services Research* 2010, **10**(1):201.
92. Greenspan JA, McMahon SA, Chebet JJ, Mpunga M, Urassa DP, Winch PJ: **Sources of community health worker motivation: a qualitative study in Morogoro Region, Tanzania**. *Human Resources for Health* 2013, **11**(1):52.
93. Creswell JW: *Research Design: Qualitative, quantitative, and mixed method approaches*. Second edition edition. Thousand Oaks, CA: Sage; 2003.
94. Kumar R, Ahmed J, Shaikh BT, Hafeez R, Hafeez A: **Job satisfaction among public health professionals working in public sector: a cross sectional study from Pakistan**. *Human Resources for Health* 2013, **11**(2).
95. **Results-Based Financing for Health** [http://www.rbfhealth.org/about]
96. Hackman JR, Oldham GR: **Motivation through the design of work: Test of a theory**. *Organizational behavior and human performance* 1976, **16**:250-279.

97. Spector PE: **Measurement of Human Service Staff Satisfaction: Development of the Job Satisfaction Survey**. *American Journal of Community Psychology* 1985, **13**(6):693-712.
98. Flesch R: **A new readability yardstick**. *Journal of Applied Psychology* 1948, **32**(3):221-233.
99. Haag-Higuchi R: **Persian word for motivation**. Edited by Dale E; 2012.
100. Gadermann AM, Guhn M, Zumbo BD: **Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide**. *Practical Assessment, Research & Evaluation* 2012, **17**(3):1-13.
101. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ: **Evaluating the Use of Exploratory Factor Analysis in Psychological Research**. *Psychological Methods* 1999, **4**(3):272-299.
102. Rajabi MZ: **Q on psychology - Kayhan recom to contact you**. Edited by Dale E. Kabul; 2012.
103. Hamilton LC: *Statistics with STATA: Updated for Version 12*. Eighth Edition edition. Boston: Brooks/Cole, Cengage Learning; 2013.
104. Muthen LK, Muthen BO: **Mplus User's Guide**. Seventh Edition edition. Los Angeles, CA: Muthén & Muthén; 2012.
105. StataCorp: **Stata Statistical Software: Release 12**. College Station, TX: StataCorp LP; 2011.
106. Bagozzi RP, Yi Y: **On the Evaluation of Structural Equation Models**. *Journal of the Academy of Marketing Science* 1988, **16**(1):74-94.
107. Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P: **An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models**. *Sociological Methods & Research* 2008, **36**(4):462-494
108. Hu L, Bentler PM: **Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives**. *Structural Equation Modeling: A Multidisciplinary Journal* 1999, **6**(1):1-55.
109. Rigdon EE: **CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling**. *Structural Equation Modeling: A Multidisciplinary Journal* 1996, **3**(4):369-379.
110. Wang J, Wang X: *Structural equation modeling: application using Mplus*. West Sussex: John Wiley & Sons Ltd; 2012.
111. Browne MW, Cudeck R: **Alternative Ways of Assessing Model Fit**. In *Testing Structural Equation Models*. Edited by Bollen K, Long J. Newbury Park, CA: Sage; 1993:136-162.
112. **Introduction to Using Mplus 5 Part 2** [http://www.ats.ucla.edu/stat/mplus/seminars/introMplus_part2/sem.htm]
113. DeVellis RF: *Scale Development: Theory and Applications*. Second Edition edition. Newbury Park: Sage; 2003.
114. Greer T, Dunlap WP, Hunter ST, Berman ME: **Skew and internal consistency**. *Journal of Applied Psychology* 2006, **91**(6):1351-1358.
115. Cortina JM: **What is Coefficient Alpha? An Examination of Theory and Applications**. *Journal of Applied Psychology* 1993, **78**(1):98-104.
116. Bentler PM: **Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability**. *Psychometrika* 2009, **74**(1):137-143.

117. Shields MA, Ward M: **Improving nurse retention in the National Health Service in England: the impact of job satisfaction on intentions to quit**. *Journal of Health Economics* 2001, **20**(5):677–701.
118. Yıldız Z, Ayhan S, Erdogmus Ş: **The impact of nurses' motivation to work, job satisfaction, and sociodemographic characteristics on intention to quit their current job: An empirical study in Turkey**. *Applied Nursing Research* 2009, **22**(2):113-118.
119. Ryan RM, Deci EL: **Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being**. *American Psychologist* 2000, **55**(1):68-78.
120. Deci EL, Connell JP, Ryan RM: **Self-Determination in a Work Organization**. *Journal of Applied Psychology* 1989, **74**(4):580-590.
121. Greco PJ, Eisenberg JM: **Changing Physicians' Practices**. *The New England Journal of Medicine* 1993, **329**(17).
122. Gagne M, Forest J: **The Study of Compensation Systems Through the Lens of Self-Determination Theory:Reconciling 35 Years of Debate**. In *Scientific Series*. Montreal: Centre interuniversitaire de recherche en analyse des organisations (CIRANO); 2011.
123. Raykov T: **Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha**. *British Journal of Mathematical and Statistical Psychology* 2008, **61**(2):275-285.
124. Sharma S, Weathers D: **Assessing generalizability of scales used in cross-national research**. *International Journal of Research in Marketing* 2003, **20**:287-295.
125. Steenkamp J-BEM, Baumgartner H: **Development and cross-cultural validation of a short form of CSI as a measure of optimum stimulation level**. *International Journal of Research in Marketing* 1995, **12**:97-104.
126. Streiner DL, Norman GR: **Health Measurement Scales: A practical guide to their development and use**. Fourth Edition edition.: Oxford Scholarship Online; 2009.
127. Zwick WR, Velicer WF: **Comparison of Five Rules for Determining the Number of Components to Retain**. *Psychological Bulletin* 1986, **99**(3):432-442.
128. World Health Organization: **Everybody's business: Strengthening health systems to improve health outcomes. WHO's framework for action**. Geneva: World Health Organization; 2007.
129. Lam CF, Gurland ST: **Self-determined work motivation predicts job outcomes, but what predicts self-determined work motivation?** *Journal of Research in Personality* 2008, **42**:1109–1115.
130. van Saane N, Sluiter JK, M. VJHA, Frings-Dresen MHW: **Reliability and validity of instruments measuring job satisfaction--a systematic review**. *Occupational Medicine* 2003, **53**(3):191-200.
131. Luboga S, Hagopian A, Ndiku J, Bancroft E, McQuide P: **Satisfaction, motivation, and intent to stay among Ugandan physicians: a survey from 18 national hospitals**. *International Journal of Health Planning and Management* 2011, **26**(1):2-17.

132. Brewer CS, Kovner CT, Greene W, Cheng Y: **Predictors of RNs' intent to work and work decisions 1 year later in a U.S. national sample**. *International Journal of Nursing Studies* 2009, **46**(7):940-956.
133. Deci EL, Ryan R, Gagne M, Leone DR, Usunov J, Kornazheva BP: **Need Satisfaction, Motivation, and Well-Being in the Work Organizations of a Former Eastern Bloc Country: A Cross-Cultural Study of Self-Determination**. *Personality and Social Psychology Bulletin* 2001, **27**(8):930-942.
134. Central Intelligence Agency: **Afghanistan**. In *The World Factbook 2013-14*. Washington, DC: Central Intelligence Agency; 2013.
135. Chen L, Evans T, Anand S, Boufford JI, Brown H, Chowdhury M, Cueto M, Dare L, Dussault G, Elzinga G *et al*: **Human resources for health: overcoming the crisis**. *The Lancet* 2004, **364**(9449):1984-1990.
136. The World Bank: **Making Services Work for Poor People**. In *World Development Report 2004*. Washington, DC: The World Bank; 2003.
137. Eijkenaar F, Emmert M, Scheppach M, Schöffski O: **Effects of pay for performance in health care: A systematic review of systematic reviews**. *Health Policy* 2013, **110**(2-3):115-130.
138. Witter S, Zulfiqur T, Javeed S, Khan A, Bari A: **Paying health workers for performance in Battagram district, Pakistan**. *Human Resources for Health* 2011, **9**(1):23.
139. Little RJ: **Regression With Missing X's: A Review**. *Journal of the American Statistical Association* 1992, **87**(420):1227-1237.
140. Downey R, King C: **Missing data in Likert ratings: A comparison of replacement methods**. *Journal of General Psychology* 1998, **125**(2):175-191.
141. Shrive FM, Stuart H, Quan H, Ghali WA: **Dealing with missing data in a multi-question depression scale: a comparison of imputation methods**. *BMC Medical Research Methodology* 2006, **6**(57):10.
142. Ministry of Public Health: **National salary policy for the BPHS facilities in Afghanistan**. Kabul, Afghanistan: Ministry of Public Health; 2011.
143. Heeringa SG, West B, Berglund PA: *Applied Survey Data Analysis*. Boca Raton: CRC Press; 2010.
144. Rabe-Hesketh S, Skrondal A, Pickles A: **Generalized multilevel structural equation modeling**. *Psychometrika* 2004, **69**(2):167–190.
145. Xue QL: **Lecture: Introduction to Structural Equations**. In *Statistics for Psychosocial Research II: Structural Models*. Baltimore, MD: Johns Hopkins School of Public Health; 2011.
146. Bollen K: *Structural Equations with Latent Variables*. New York, NY: A Wiley-Interscience Publication; 1989.
147. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB: **Analysis of clinical trials by treatment actually received: is it really an option?** *Statistics in Medicine* 1991, **10**(10):1595-1605.
148. Heitjan DF: **Causal Inference in a Clinical Trial: A Comparative Example**. *Controlled Clinical Trials* 1999, **20**:309–318.
149. Wright CC, Sim J: **Intention-to-treat approach to data from randomized controlled trials: A sensitivity analysis**. *Journal of Clinical Epidemiology* 2003, **56**:833-842.

150. Jo B, Asparouhov T, Muthen BO, Ialongo NS, Brown CH: **Cluster Randomized Trials With Treatment Noncompliance**. *Psychological Methods* 2008, **13**(1):1-18.
151. Little RJ, Yau LHY: **Statistical Techniques for Analyzing Data From Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model**. *Psychological Methods* 1998, **3**(2):147-159.
152. Little RJ, Rubin DB: **Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes: Concepts and Analytical Approaches**. *Annu Rev Public Health* 2000, **21**:121-145.
153. Jo B: **Estimation of Intervention Effects with Noncompliance: Alternative Model Specifications**. *Journal of Educational and Behavioral Statistics* 2002, **27**(4):385-4009.
154. Stuart EA, Perry DF, Le H-N, Ialongo NS: **Estimating Intervention Effects of Prevention Programs: Accounting for Noncompliance**. *Prevention Science* 2008, **2008**(9):288-298.
155. UCLA SCG: **Regression Diagnostics**. In *Regression with Stata*. UCLA: Statistical Consulting Group.
156. Jo B, Ginexi EM, Ialongo NS: **Handling Missing Data in Randomized Experiments with Noncompliance**. *Prevention Science* 2010, **11**(4):384-396.
157. Rothman KJ, Greenland S, Lash TL: *Modern Epidemiology*. Third edition edition. Baltimore, MD: Lippincott Williams & Wilkins; 2008.
158. Oshang N: **Re: RBF - question**. Kabul; 2013.
159. Rabe-Hesketh S, Skrondal A: *Multilevel and Longitudinal Modeling Using Stata*. Third edition edition. College Station, Texas: STATA Press; 2012.
160. Jo B: **Re: Your study on Handling Missing Data in Randomized Experiments with NonCompliance**. Stanford; 2013.
161. Kalk A, Paul FA, Grabosch E: **'Paying for performance' in Rwanda: does it pay off?** *Tropical Medicine and International Health* 2010, **15**(2):182-190.
162. Ministry of Public Health: **National Salary Policy For Non-Governmental Organizations and Ministry of Health Strengthening Mechanism Working in the Afghan Health Sector**. Kabul: Ministry of Public Health; 2005.
163. Fritsche G, Vergeer P: **Performance-Based Financing Drill Down**. In *Results-Based Financing Workshop for Bank Staff: November 2, 2010 2010*. The World Bank; 2010.
164. Oshang N: **RE: Three quick questions on RBF**. Kabul; 2013.
165. Bhatnagar A: **Evaluation of the Results-based financing in health pilot in Nigeria.** . Unpublished.
166. Oshang N: **Reasons for negative effect**. Kabul; 2013.
167. Johns Hopkins Bloomberg School of Public Health, Ministry of Public Health of the Islamic Republic of Afghanistan, Indian Institute of Health Management Research: **Afghanistan Basic Package of Health Services Balanced Scorecard: National Report 2009-10**. Kabul; 2011.
168. **Factor scores** [http://www.statmodel.com/faq.shtml]
169. **How MPLUS Computes Factor Scores** [http://www.statmodel.com/discussion/messages/9/3778.html]

170. Skrondal A, Laake P: **Regression among factor scores**. *Psychometrika* 2001, **66**(4):563-576.
171. DiStefano C, Zhu M, Mindrila D: **Understanding and Using Factor Scores: Considerations for the Applied Researcher**. *Practical Assessment, Research & Evaluation* 2009, **14**(20):1-11.
172. Enders CK, Tofighi D: **Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue**. *Psychological Methods* 2007, **12**(2):121-138.
173. Hausman JA: **Specification tests in econometrics**. *Econometrica* 1978, **46**(6):1251-1271.
174. Donabedian A: *The Definition of Quality and Approaches to Its Assessment. Explorations in Quality Assessment and Monitoring*. *Volume 1*. Ann Arbor, MI: Health Administration Press; 1980.
175. Beaulieu ND, Horrigan DR: **Putting Smart Money to Work for Quality Improvement**. *Health Services Research* 2005, **40**(5p1):1318-1334.
176. Leonard KL, Masatu MC, Vialou A: **Getting Doctors to Do Their Best: The Roles of Ability and Motivation in Health Care Quality**. *Journal of Human Resources* 2007, **XLII**:682-700.
177. Das J, Gertler PJ: **Variations In Practice Quality In Five Low-Income Countries: A Conceptual Overview**. *Health Affairs* 2007, **26**(3):w296-w309.
178. Serneels P, Lindelow M, Montalvo JG, Barr A: **For public service or money: understanding geographical imbalances in the health workforce**. *Health Policy and Planning* 2007, **22**(3):128-138.

Curriculum Vita

Elina M. Dale
Date and place of birth: April 2, 1978, Przhevalsk
Email: emanzhie@jhsph.edu

EDUCATION

2014: **The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD**
Doctor of Philosophy, Department of International Health, Health Systems Program
Advisor: David H. Peters
Dissertation topic: Performance-Based Payments, Provider Motivation and Quality of Care in Afghanistan

Awards: Open Society Institute Global Supplementary Grant Program 2010 – 2011, 2011 – 2012

2004: **Fletcher School of Law and Diplomacy, Tufts University, Medford, MA**
MA, Development Economics and Southwest Asia & Islamic Civilizations, GPA: 3.76/4.0

Awards: American Association of University Women International Fellowship 2001-2002

2001: **Central European University, Budapest, Hungary**
MA, Gender & Culture Studies, GPA: 3.72/4.0

2000: **American University in Central Asia, Bishkek, Kyrgyzstan**
BA, International Relations, GPA: 3.83/4.3

SELECTED PUBLICATIONS

- Giuffrida A, Jakab M, Dale EM. *Toward Universal Coverage in Health: The Case of the State Guaranteed Benefit Package of the Kyrgyz Republic.* Universal Health Coverage Studies Series: No. 17. Washington, DC: World Bank, 2013.

- Jakab M, Manjieva E. In Gottret P, Schieber G, eds. *Good Practice in Health Care Financing: Lessons from Reforms in Low- and Middle-Income Countries*. Washington, DC: World Bank, 2006

- Ibraimova A, Akkazieva B, Ibraimov A, Manzhieva E, Rechel B . Kyrgyzstan: Health system review. *Health Systems in Transition*. European Observatory on Health Systems and Policies, 2011

- Манжиева Э, Кожокеев К, Мурзалиева Г . *Оценка программы «Депозит врача»* (Evaluation of Financial Incentives to Retain Doctors in Rural Areas and Small Towns). Policy Research Paper N 52. Bishkek: WHO/DFID Health Policy Analysis Project & CHSD, 2008

- Manjieva E, Narmanbetov U, Kadyrova N, Jakab M. *Analysis of the Medium-Term Financial Sustainability of the State Guaranteed Benefits Package*. Policy Research Paper N 43. Bishkek: WHO/DFID Health Policy Analysis Project, MHIF and CHSD, 2007

- Contributed to World Bank studies, incl. Human Development Sector Unit, Europe & Central Asia Region. *Operationalizing the Health and Education Millennium Development Goals in Central Asia*. Washington, DC: World Bank, 2005

LANGUAGE & STATISTICAL SOFTWARE SKILLS

Languages: Russian (Native), English (Fluent), Norwegian (Intermediate in reading & writing)
Statistical software: STATA, MPlus

EMPLOYMENT

World Bank, Bishkek Kyrgyzstan 05.2012 – 06.2013
Consultant, Programmatic Public Expenditure Review – Health Component

- Reviewed the state treasury reports, tracking health sector budget and expenditures over the past ten years
- Contributed to a policy note on the fiscal rules governing the health SWAP
- Contributed to a note reviewing the State Guarantees Benefit Package, focusing on the coverage of the poor
- Draft a report assessing public expenditure on health from a fiscal space perspective, focusing on efficiency measures to reduce the funding gap of the SGBP

Johns Hopkins School of Public Health, Baltimore, MD 02.2011 – 05.2013
Research Assistant, Health Systems Program

- Contributed to revision & pilot testing of health facility assessment tools for the Balance Score Card under the National Health System Performance Assessment in Afghanistan
- Analyzed data from the baseline household survey for the evaluation of the Results-Based Financing Project in Afghanistan and contributed to writing of the report
- Analyzed data & co-authored the report on the national health system performance assessment in Afghanistan, focusing on equity of health services utilization

Evaluation of Norwegian ESTHER Program, Oslo, Norway 11.2011 – 12.2011
Client: Norwegian Development Agency (NORAD)

- Contributed to writing of the report, specifically literature review of twinning programs & staff exchanges between health care organizations
- Conducted telephone interviews with participants of the ESTHER Program - Norway

Johns Hopkins School of Public Health, Baltimore, MD 09.2010 – 02.2011
Project Assistant, Road Safety in 10 Countries (RS-10), the Russian Federation

- Assisted in setting up contracts & communicating with institutional partners in the Russian Federation

- Analyzed technical documents & data in Russian, and contributed to writing of the baseline report.

Redd Barna (Save the Children), Oslo, Norway 03.2010 – 12.2010
Program Officer, EVERY ONE Campaign

- Provided technical feedback to SC in Nepal & Uganda during the development of the National Campaign Plans, including overall structure, specific activities, logical frameworks and budgets
- Contributed to development of the funding proposal for the Campaign activities in Afghanistan, Ethiopia, India, Nepal and Uganda to the Norwegian Ministry of Foreign Affairs
- Contributed to the work of the Global M&E Task Force, including the development of global & initiative indicators, reporting tool, & technical guidance for the field offices

Johns Hopkins School of Public Health, Kabul, Afghanistan 06.2010 – 09.2010
Technical Advisor, Results-Based Financing Intervention in Afghanistan

- Developed household survey instrument & contributed to development of the field guidance manual for the survey teams
- Drafted the Impact Evaluation Design Note for RBF, including sampling approach & sample size estimates
- Participated in meetings with local and international stakeholders, including Afghan Ministry of Public Health and the World Bank

World Health Organization, Bishkek, Kyrgyzstan 10.2009 – 11.2009
Consultant, Preparatory Mission on Results-Based Financing in Health

- Produced district-by-district map of projects in MCH based on review of different project documents and interviews with development partners
- Developed Results Chain Frameworks for the three priority areas - maternal health, neonatal health & child health in collaboration with the MOH & development partners
- Organized and led a series of meetings with local and international stakeholders to develop payment indicators and a budget necessary for improvement of existing data collection system

Johns Hopkins School of Public Health, Baltimore, MD 04.2009 – 08.2009
Research Assistant, Health Systems Program

- Analyzed household survey data on reproductive & child health in Afghanistan
- Contributed to the review of the household survey for Zambia and development of the manual on indicators under the Results-Based Financing in Health Program.

World Health Organization, Dushanbe, Tajikistan 08.2007 – 12.2007
Consultant, GAVI health system strengthening (HSS) support country application

- Provided technical input to the Working Group (WG) in mapping all donor activities in the health sector and identifying key bottlenecks
- Conducted a health system needs assessment
- Based on the WG discussions and proposals, wrote up description of components for the proposal

- Provided technical input to the WG in prioritization of activities & geographical targeting based on the needs assessment, and in developing monitoring framework and indicators

World Health Organization, Yerevan, Armenia 06.2007 – 07.2007
Consultant, GAVI health system strengthening (HSS) support country application

- Conducted a health system needs assessment
- Provided technical input to the WG in linking identified bottlenecks, proposal objectives and activities

Ministry of Health of the Kyrgyz Republic, Bishkek, Kyrgyzstan 05.2006 – 05.2008
Policy Analyst, WHO/DFID Health Policy Analysis Project

- Supervised the preparation of quarterly Financial Management Reports by the Ministry staff, including comparison of original budget with actual expenditures & sources of discrepancies
- Participated in annual technical reviews of the National Health Reform Program, contributing to the Joint Annual Review notes
- Coordinated follow-up actions & dialogue with the Ministry of Finance to ensure fulfillment of public financing obligations under the National Health Reform Program
- Contributed to the development of the first regional course on *Accelerating Progress towards the Health MDGs for Central Asia and the Caucasus* for health policy makers, facility managers & NGO representatives

World Bank Country Office, Bishkek, Kyrgyzstan 04.2004 – 05.2006
Junior Professional Associate (JPA), Human Development Unit

- Contributed to preparation of diverse operational products, including project operational manuals & project appraisal documents for Health SWAp, Rural Education Project & Central Asia AIDS Control Project
- Maintained dialogue with the Government of the Kyrgyz Republic at the technical expert and ministerial levels to learn about new policy initiatives, inform Bank management about them, and design appropriate & timely recommendations
- Monitored public financing of the health sector on a regular basis and contributed to annual country assessment, sector briefings, and public expenditure reviews (PERs)
- Provided technical assistance to the Ministries of Health and Education in developing medium-term expenditure frameworks (MTEF) and program-based budgets in accordance with sector priorities
- Provided support to project implementation units during procurement process, including developing terms of references, evaluating technical and financial proposals, reviewing and following up on progress reports
- Assisted the World Bank and DFID senior financial management specialists in development and follow-up of the fiduciary risks mitigations measures under the Health SWAp

World Bank Institute, Washington, DC 05.2003 – 04.2004
Evaluation Consultant, Evaluation Group

- Conducted statistical analyses of survey data and wrote a chapter for the impact evaluation report on the World Bank Institute programs in the Russian Federation
- Conducted interviews with field staff and edited the regional impact evaluation study of WBI programs in East Asia and the Pacific

PROFESSIONAL ASSOCIATIONS AND CONFERENCES

European Conference on Health Economics, Helsinki, Finland 01.2010
Review and scoring of 30 abstracts

Stortingsvalget 2009, Oslo, Norway 09.2009
Translator for international election observers

European Conference on Health Economics, Rome, Italy 01.2008
Review and scoring of 40 abstracts

6th World Congress on Health Economics, Copenhagen, Denmark 07.2007
Presented a paper on "Evaluating Effects of Health Care Reforms on Coverage in the Kyrgyz Republic 1991-2006"

International Health Economics Association 03.2007 – 03.2008
Member

Harvard International Development Conference, Cambridge, MA 01.2002 – 04.2002
Member of the Conference Steering Committee responsible for technical content of the Human Development Panel and fundraising for the Conference

PROFESSIONAL TRAINING

Poverty, Equity and Health Systems 10.2006
The World Bank Institute, Hungary

Accelerating Progress Towards the Health MDGs 03.2006
The World Bank Institute, Washington, DC

World Bank Procurement Workshop 02.2006
The World Bank, Kyrgyzstan

Public Health Leadership Programme 07.2005
Imperial College London